NIST Special Publication 1276v3

NIST Conference Papers
Fiscal Year 2019
Volume 3:
Information Technology Laboratory
Material Measurement Laboratory
Physical Measurement Laboratory

Compiled and edited by:
Information Services Office

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1276v3

National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Special Publication 1276v3

NIST Conference Papers Fiscal Year 2019

Volume 3:
Information Technology Laboratory
Material Measurement Laboratory
Physical Measurement Laboratory

Compiled and edited by:
Resources, Access, and Data Team
Information Services Office

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1276v3

April 2022

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Special Publication 1276v3
Natl. Inst. Stand. Technol. Spec. Publ. 1276v3, 807 pages (April 2022)
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1276v3

Foreword

NIST is committed to the idea that results of federally funded research are a valuable national resource and a strategic asset. To the extent feasible and consistent with law, agency mission, resource constraints, and U.S. national, homeland, and economic security, NIST will promote the deposit of scientific data arising from unclassified research and programs, funded wholly or in part by NIST, except for Standard Reference Data, free of charge in publicly accessible databases. Subject to the same conditions and constraints listed above, NIST also intends to make freely available to the public, in publicly accessible repositories, all peer-reviewed scholarly publications arising from unclassified research and programs funded wholly or in part by NIST.

This Special Publication represents the work of researchers at professional conferences, as reported in Fiscal Year 2019.

More information on public access to NIST research is available at https://www.nist.gov/open.

Key words

NIST conference papers; NIST research; public access to NIST research.

Table of Contents

Standard Radio Communication Test Method for Mobile Ground Robots

Galen Koepke*. David Hooper, John M. Andrews**, and Kate A. Remley***

**National Institute of Standards and Technology, Boulder, CO 80305*
USA (Tel: 303-497-5766; e-mail: koepke, remley@ boulder.nist.gov).
***Unmanned Systems Branch of the Space and Naval Warfare Systems Center Pacific, San Diego, CA 92152, USA (e-mail: dhooper@spawar. navy.mil, john.andrews@navy.mil)*

Abstract: This paper describes a proposed standard test method for evaluating the performance of wireless communications links used for control and telemetry in mobile ground robots. Range performance metrics are determined under open line-of-sight (LOS) and simple obstructed non-line-of-sight (NLOS) conditions. The method was demonstrated as part of the annual Response Robot Evaluation Exercise at the Texas A&M University Riverside Campus during November 2008, where it was used to compare range performance of several advanced radio systems.

Keywords: Radio communication, robots, standards, teleoperation, testing, wireless.

1. INTRODUCTION

Teleoperated robotic vehicles are rapidly emerging as an invaluable tool for the military as well as civil emergency response organizations. Remote operation offers clear advantages in applications such as hazardous materials response, mine safety, urban search and rescue, and reconnaissance. The stand-off provided by telerobotics can serve to reduce human exposure to dangerous situations and environments while also providing opportunities to extend the users' tactical reach into otherwise inaccessible spaces.

A critical performance parameter influencing, if not determining, overall system capability is the communication link between the user and the robot. This link is typically duplex, serving as the means of issuing remote commands to the robot while also providing one or more channels for transmitting sensor data, usually including live video images, back to the controlling user. The communication link may be wired but is more commonly wireless to facilitate enhanced mobility; *i.e.*, movement without the encumbrance of a tether. Range performance of wireless teleoperated robots varies between systems as a function of several design parameters including: radio frequency, power, antenna gain, antenna height, etc. Environmental conditions strongly affect range performance as well, because wireless systems are subject to operational range limitations due to path attenuation, latency, and interference arising from other wireless users, ambient noise, and multipath effects. Depending on system specifications and environmental conditions, the maximum effective range of commercially available teleoperated ground robots generally varies from tens of meters to several kilometers.

Currently there are no standard test methods for obtaining and evaluating performance metrics for the wireless telemetry and control systems of ground robots. The U.S. Department of Homeland Security has tasked the National Institute of Standards and Technology (NIST) to address this need as part of its comprehensive effort to develop standards related to the testing, evaluation, and certification of robotic technologies [1]-[3]. NIST is actively working to develop standards for mobility, perception, navigation, endurance, and human factors, as well as communications. The NIST program places specific emphasis on standards appropriate for urban search-and-rescue (US&R) scenarios, but its practices are widely applicable to other operational domains employing mobile robots.

This paper describes two of the preliminary standard test methods for communications and their recent use to evaluate and compare range performance metrics for several radio technology alternatives on a small ground robot. The two methods include a line-of-sight range test and a non-line-of-sight range test.

2. PERFORMANCE TEST METHODS

Test methods to ensure that a given robot will meet the needs of a given response agency must be repeatable, reproducible, and must isolate various characteristics of robot performance. When a number of potential impediments to robot performance are present simultaneously, it is impossible to assess a specific characteristic of the wireless system. This was illustrated clearly in [2], which described initial field tests where various robots were exposed to preliminary test methods. In [2], robots were unable to complete line-of-sight and non-line-of-sight tests because of radio interference from other robots nearby. As a result, the test methods discussed below were developed to test specific characteristics of robot performance, in isolation from other effects. These methods are intended to test only the specific attributes of (1) loss of radio communication in a line-of-sight condition as the robot moves down range and (2) loss of radio communication as the robot moves behind an obstacle.

Remley, Catherine; Koepke, Galen.
"Standard Radio Communication Test Method for Mobile Ground Robots."
Paper presented at 2nd International ICST Conference on Robot Communication and Coordination, ROBOCOMM 2009, Odense, Denmark.
March 31, 2009 - April 2, 2009.

In order to isolate these attributes, the performance test methods are necessarily abstract, because no real-world environment will present such isolated conditions. It is also necessary to conduct these tests in a facility where they may be repeated and reproduced as often as necessary. The test facility chosen to meet the above criteria of isolation and reproducibility is an airstrip at the Texas A&M University Riverside Campus. The preliminary exercise of [2] and other response robot field exercises has led to the standardized test methods described below. These test methods are the first two in a series of tests that will be needed to fully characterize robot performance. Such methods could be useful in characterizing other types of robots, as well as other types of field-deployable wireless communication equipment.

2.1 Line-of-sight (LOS) range test method

This test effectively replicates the simplest propagation environment likely to be encountered by radio-linked robotic systems on the ground. The test scenario is intended to evaluate system communication range under conditions of unobstructed visual line-of-sight between the robotic vehicle and the base-station antenna. This will typically involve two antennas separated by a distance (d), though the scenario could accommodate multiple nodes, such as in the case of a robot that carries and deploys repeaters to extend its communication range [4].

Radio propagation under these conditions can be described by the two-ray model over plane earth [5]. This model assumes only two signals arriving at the receive antenna: the direct (shortest line-of-sight path) signal between the two antennas, and one signal that reflects (bounces) off the earth before it arrives at the receiving antenna. The reflected signal typically has a different amplitude and phase compared to those of the direct signal. The method assumes relatively short separation distances ($d < 2.0$ km), hence the curvature of the earth can be ignored. A minimum of 50 meters on each side of the test path (100 meters total) must be clear of any obstacles or reflecting objects in order to minimize multipath effects.

Test points are established at regular intervals (Fig.1). The reference location is at 50 m. The second location is at 100 m, with subsequent locations at every 100 meters. Test locations at shorter intervals (*e.g.*, 50 m) may also be used if needed.

Test procedures: At each test point the mobile robot is made to perform a full system evaluation while facing in each of the four compass directions relative to the direction of travel. This is accomplished while maneuvering through a tight figure-eight course and selecting four test stations such that the robot is facing in each direction (Fig 2). Performance testing at each test point serves to validate function of both the control and data (video) links while removing orientation bias in cases where either the radiation pattern of the on-board antenna system is not perfectly omnidirectional or the antenna is obstructed by an asymmetric part of the robot platform, for example, a manipulator arm.

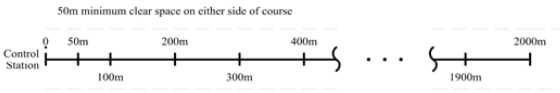


Fig. 1. Line-of-sight course layout showing spacing between test points

The LOS range test sequence is as follows:

Step 1: Perform a full system(s) evaluation in each of four directions (*e.g.*, read visual charts, read on-board sensors, operate manipulator, etc.) at a reference point 50 meters from the control point. This is the reference test and is used to determine the best-case performance for the communications systems.

Step 2: Proceed to the next test point location (*e.g.*, 100 m, 200 m, etc.) and perform a full system(s) evaluation in each of four directions.

Step 3: Repeat step 2 until any system operation becomes unacceptable (as defined for the particular system) or the end of the course is reached at 2000 meters.

2.2 Non-line-of-sight (NLOS) range test method

The NLOS test is intended to evaluate the effectiveness of the robotic system's radio communication link under conditions where the visual line-of-sight between the robot and base station is completely occluded by a physical object. In this case, neither the base station nor the robot's onboard antenna receives a direct signal but instead they receive one or more significantly attenuated signals due to diffraction around the obstacle. The course is located at least 50 m from any other obstacles in order to minimize reflection and/or scattering from a nearby object.

This is intended to serve as a simplified replication of a scenario where the robot is used to explore an object or incident beyond or out of the visual field of the operator. The test studies path loss that is due to attenuation rather than scattering.

The NLOS test includes two sections: a 500 meter line-of-sight portion with characteristics similar to the LOS test, after which the robot makes a 90 degree turn behind a tall, broad obstruction such as a building and continues the test (Fig. 3). Both tests are executed in a manner identical to the LOS test in that the robot is made to maneuver through a figure-eight while confirming visual acuity four times at regularly spaced test points. However the test point spacing is shortened for the non-line-of-sight portion of the test.

Remley, Catherine; Koepke, Galen.
"Standard Radio Communication Test Method for Mobile Ground Robots."
Paper presented at 2nd International ICST Conference on Robot Communication and Coordination, ROBOCOMM 2009, Odense, Denmark.
March 31, 2009 - April 2, 2009.

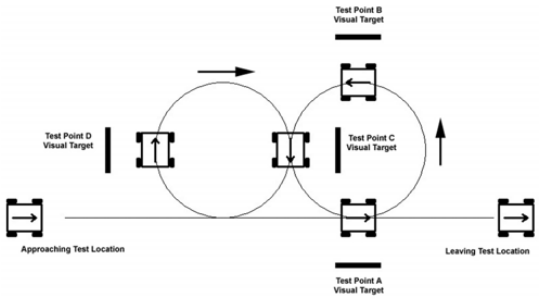


Fig. 2. Position sequence at each test point. Robot traverses a figure-eight and executes four visual acuity tests while halted in front of visual targets. The test ensures that the robot will operate in all four compass directions by requiring it to face the pre-oriented targets.

Obstacle (building) requirements: The building or obstacle should provide complete shadowing of the radio-frequency signal for the path taken by the robot. It should be constructed of materials that prevent easy penetration of the signal and should be both tall and broad enough to prevent signals from reaching the reverse side of the structure, for example, a metallic wall. The standard test course may be assembled from 12.2 meter long (40 foot) International Organization for Standardization (ISO) shipping containers stacked three high and at least two wide such that the resulting obstacle measures 7.8 meters (25.5 feet) tall by 24.4 meters (80 feet) wide. The electrically conductive steel construction of shipping containers makes for an excellent radio-wave barrier for most frequencies of operation currently used by robotics manufacturers. To ensure no energy leakage, metal foil and tape should be used to cover any gaps between containers.

Similar test courses may be used to provide an estimate of robot performance for this test method. For instance, a large commercial multi-story concrete and steel structure covering a large portion of a city block would be acceptable, while a wood-frame single family home is unlikely to provide adequate attenuation or shadowing. The results may be complicated by an obstruction that is too small to sufficiently block the direct path signal, or a highly reflective environment, such as an urban canyon, where signals are channeled along side streets by tall structures; such conditions should be avoided.

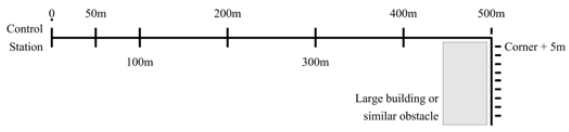


Fig. 3. Non-line-of-sight test course configuration showing LOS portion leading to a 90 degree turn behind an obstacle that blocks the direct radio propagation path. Test points are set up at regular intervals.

There are two separate and distinct propagation profiles that result from the two parts of the test path. The LOS portion is expected to follow the two-ray model, as before. The signal attenuation associated with the NLOS section will increase rapidly after the mobile robot moves behind the obstacle. The weakened NLOS signal can be attributed to a combination of indirect paths, including diffraction around the obstacle as well as small amounts of scattering and reflection from nearby objects, although care is taken to minimize the latter.

Test procedures: At each test point the mobile robot is made to maneuver through a figure-eight and perform a full system evaluation while facing in each of the four compass directions, just as in the LOS test. Some robot systems may not have the ability to communicate over the full length of the 500 meter LOS portion of the test course, making it impossible to reach the NLOS part of the course. In this case the system start point should be set at a distance from the obstruction equal to not more than one half of the maximum achievable LOS range, as determined by the LOS range test described above.

The NLOS range test sequence is as follows:

Step 1: Initiate the LOS section, conducting full system(s) evaluation (*e.g.*, read visual charts, read on-board sensors, operate manipulator, etc.) at 50 meters from control point.

Step 2: Proceed to the next test locations iteratively (100, 200, 300, 400, and 500 m) and perform a full system(s) evaluation.

Step 3: Repeat (2) until any system operation becomes unacceptable (as defined for a particular system) or until the system reaches the far corner of the obstacle at 500 meters.

Step 4: Maneuver the robot 90 degrees so that it is behind the obstacle. This is the start of the NLOS section. The first test station is located 5 m from the corner of the obstacle (building). The test station and the robot path should be positioned as close as possible to the obstacle, with one of the test charts directly on the surface of the structure. Perform a full system(s) evaluation (*e.g.*, read visual charts, read on-board sensors, operate manipulator, etc) at the 5 m test location.

Step 5: Proceed to next test point location (at 5 m or 10 m intervals from the corner).

Step 6: Perform a full system(s) evaluation.

Step 7: Repeat Step 6 until any system operation becomes unacceptable (as defined for a particular system) or until reaching the far corner of the obstacle.

Remley, Catherine; Koepke, Galen.
"Standard Radio Communication Test Method for Mobile Ground Robots."
Paper presented at 2nd International ICST Conference on Robot Communication and Coordination, ROBOCOMM 2009, Odense, Denmark.
March 31, 2009 - April 2, 2009.

Fig. 4. Line-of-sight course set up over a 1 km distance on an airstrip with orange safety cones marking test points.

3. CONDUCTING THE TESTS

NLOS and LOS range performance tests, using the standard methods described above, were conducted as part of the NIST sponsored Response Robot Evaluation Exercise held 17-20 November 2008 in College Station, Texas. The communication test portion of the event was conducted at the Texas A&M University Riverside Campus, just west of Bryan, Texas. The event was facilitated by the Texas A&M Texas Engineering Extension Service (TEEX).

The Riverside Campus location was originally built as an Army Airfield. The former airfield's runways, no longer in use, serve as an excellent location to conduct radio range testing. The hard concrete surfaces of the airstrips provide a uniform reproducible environment, the terrain is flat, and the facility provides kilometers of maneuvering space that is relatively free of existing obstructions. The LOS test course was established on one runway, the NLOS course on an adjoining runway.

Fig. 5. Example test station for both the LOS and NLOS methods. Three visual acuity charts are set up around the figure-eight maneuver path.

Fig. 6. Example of a visual acuity chart used at each test station for both the LOS and NLOS methods

The set-up for LOS testing is shown in Fig. 4. Orange safety cones were used to mark test stations along the course (Fig. 5), visual acuity charts were mounted in orange safety triangles, as shown in Fig. 6. The length of the LOS course was limited to one kilometer, rather than two, with a reference point established at 50 m and subsequent test points starting at 100 m and continuing at 100 m intervals as per the method guidelines.

The NLOS course layout is shown in Figs. 7-8. Six steel ISO shipping containers were set up to serve as the obstruction. TEEX employees used a crane to stack the boxes two wide (24 m) and three high (7.8 m). The container stack was located at a 500 m distance from the start (control) point of the NLOS course. A reference point was set at a 50 m distance, with subsequent test points emplaced starting at the 100 m mark and continuing at 100 m intervals up to 500 m, collinear with the near edge of the obstacle.

Fig. 7. NLOS test course as viewed from near the start point looking toward the container stack 500 m away.

Remley, Catherine; Koepke, Galen.
"Standard Radio Communication Test Method for Mobile Ground Robots."
Paper presented at 2nd International ICST Conference on Robot Communication and Coordination, ROBOCOMM 2009, Odense, Denmark.
March 31, 2009 - April 2, 2009.

Fig. 8. The far side of the container stack showing placement of the final (500 m) figure-eight LOS test station and subsequent NLOS test points located along the back wall.

Prior to the start of each test run, a spectrum analyzer was used to identify and document ambient radio-frequency noise and establish baseline noise conditions. These provided a means of recognizing environmental changes that could influence performance. As the tests were conducted in a populated, albeit semirural area, some level of background activity was expected with the potential for interference within commercial and unlicensed portions of the spectrum.

The LOS and NLOS test courses were used to evaluate range performance on six unique radio systems. Frequencies ranged from 225 MHz to 5.9 GHz. Each radio system was integrated onto an iRobot PackBot EOD robot [6, 7] which served as the mobile platform for each test. Each PackBot mounted radio was maneuvered through both the LOS and NLOS test lanes to determine the maximum effective communication range for the system. Each run was replicated at least twice as an initial means of validating reproducibility.

Specific details on the radio systems tested and a comparative evaluation of their performance based in part on the methods described in this paper are to be published separately.

4. DISCUSSION

4.1 Precision and Bias

The communication test methods should be evaluated for reproducibility each time they are used. During the Response Robot Evaluation test event, the maximum achievable range determined for a given system varied somewhat from run to run. For example, two of the six tested systems were able to successfully complete the full length of the NLOS test course on one run, but failed at a point significantly short of the final test point during their next run.

Several factors can lead to variance in testing. One variable that was found to introduce bias is radio interference arising from outside transmissions in the vicinity of the test site. Ambient radio interference and noise levels are often beyond the control of the test manager. To mitigate this bias, a survey of background spectral emissions using a spectrum analyzer should be carried out prior to, or preferably, during each test. This is particularly important to help ensure accurate comparison when evaluating systems operating at different frequencies, as they may be unequally affected depending on the spectral characteristics of the interfering noise.

During the Response Robot Evaluation event, measurement and analysis of background noise led to the discovery of significant 400 MHz noise that may have limited the performance of some of the tested radio systems. Knowledge of the background noise characteristics will also allow for more informed comparisons when evaluating performance measurements gathered from different locations or at different times. Nevertheless, if a wireless system passes the test method in the presence of interference, it should not affect the rating of that system.

Other variables that may introduce bias into the wireless communication test methods are related to the environmental condition under which the tests were carried out. Atmospheric conditions, solar activity, precipitation, temperature, and humidity can all affect the results. The composition of the soil and ground at a given site can be expected to influence performance results as well. We suspect that the environmental conditions produce test results outside the standard deviation of the reproducibility of the method, so that the operator may return to the test site for additional testing when the conditions more closely match those specified in the reproducibility test.

4.2 Use of Additional Indicators

The test methods have been developed with the end user in mind; they are specifically designed to provide a go/no-go performance evaluation for a given system in a simulated environment. The method may find additional utility among developers by introducing the continuous measurement of key performance indicators such as latency, data error rate, bandwidth loss, etc. This could be used to identify specific failure mechanisms and lead to engineering changes resulting in extended range performance.

5. CONCLUSION

The development of application-specific robot standards and repeatable performance testing methods with objective metrics will accelerate the development and deployment of mobile robotic tools for the user community. Improved understanding of specific robotic capabilities and limitations will enhance the effectiveness of operator teams while serving to reduce the risks and uncertainty associated with operational use.

Adequate performance using these test methods will not ensure successful operation in all operational environments, due to possible unforeseen extreme or unusual communications difficulties present in some radio environments. Rather, these tests are intended to provide a common ground for comparison of technologies against a reasonable simulation of relevant environments and to

Remley, Catherine; Koepke, Galen.
"Standard Radio Communication Test Method for Mobile Ground Robots."
Paper presented at 2nd International ICST Conference on Robot Communication and Coordination, ROBOCOMM 2009, Odense, Denmark.
March 31, 2009 - April 2, 2009.

provide quantitative performance data to user organizations to aid in choosing appropriate systems.

The practice described in this paper provides a method for quantifying the maximum effective range performance for on-the-ground communications systems of mobile robots. NIST and its partners, including the U.S. Department of Homeland Security (DHS) and the American Society for Testing Materials (ASTM), will continue to utilize the results and lessons learned from ongoing use of the test methods, as described here, to improve and refine the protocol. This will ultimately lead to recognized and accepted standards that provide a consistent way of evaluating and comparing system performance within the growing field of available robotic platforms.

ACKNOWLEDGEMENT

The authors thank the staff of Texas A and M University of Texas Engineering Extension Service (TEEX) for their support in setting-up and facilitating test scenarios during the 2008 Response Robot Evaluation Event. The U.S. Department of Homeland Security Office of Standards supplied funding for this work.

REFERENCES

[1] Elena Messina, Adam Jacoff, Jean Scholtz, Craig Schlenoff, Hui-Min Huang, Alan Lytle, and John Blitch, Preliminary Report: "Statement of Requirements for Urban Search and Rescue Robot Performance Standards", [Online]. Available: http://www.isd.mel.nist.gov

[2] Kate A. Remley, Galen Koepke, Elena Messina, Adam Jacoff, and George Hough, Standards Development for Wireless Communications for Urban Search and Rescue Robots, Proc. ISART 2007, Boulder, CO, March 2007, pp. 79-82, [Online] Available: http://www.boulder.nist.gov/div818/81802/MetrologyForWirelessSys/pubs/R11_ISART07_Remley.pdf

[3] Elena Messina and Adam Jacoff, Performance Standards for Urban Search and Rescue Robots, Proceedings of the SPIE Defense and Security Symposium, Orlando, FL, April 17-21, 2006, [Online] Available: http://www.isd.mel.nist.gov/documents/messina/6230-77.pdf

[4] Narek Pezeshkian, Hoa G. Nguyen, and Aaron Burmeister, Unmanned Ground Vehicle Radio Relay Deployment System for Non-Line-of-Sight Operations, 13th IASTED International Conference on Robotics & Applications, Würzburg, Germany, August 29-31, 2007, [Online] Available: http://www.spawar.navy.mil/robots/pubs/IASTED_ADCR_2007.pdf

[5] Ramakrishna Janaswamy, "Radio Propagation and Smart Antennas for Wireless Communications," Kluwer Academic Publishers, 2001, eq. 2.60, pg 36.

[6] Identification of commercial products does not imply endorsement by the National Institute of Standards and Technology. Other products may work as well or better.

[7] Information available at: http://irobot.com

Remley, Catherine; Koepke, Galen.
"Standard Radio Communication Test Method for Mobile Ground Robots."
Paper presented at 2nd International ICST Conference on Robot Communication and Coordination, ROBOCOMM 2009, Odense, Denmark.
March 31, 2009 - April 2, 2009.

The Benefits of Network Coding over a Wireless Backbone

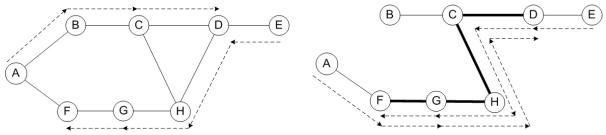
Hui Guo[†], Yi Qian[†], Kejie Lu[‡] and Nader Moayeri[†]
[†]National Institute of Standards and Technology
Gaithersburg, MD 20899, USA
[‡]University of Puerto Rico
Mayaguez, PR 00681, USA

***Abstract*— Network coding is a promising technology that can effectively improve the efficiency and capacity of multi-hop wireless networks by exploiting the broadcast nature of the wireless medium. However, current packet routing schemes do not take advantage of the network coding, and the benefits of network coding have not been fully utilized. To improve the performance gain of network coding, in this paper, we apply network coding over a wireless backbone and investigate the performance of this approach from a theoretical perspective. Our analysis shows that, compared to network coding over ad hoc networks with traditional routing schemes, network coding over the backbone structure exhibits significant advantages. This is because all packets are transmitted over the constructed backbone with pre-specified routes, and consequently the opportunity for coding packets at intermediate nodes can be substantially improved. To further enhance the performance, we also present an optimized link scheduling protocol for network coding over a wireless backbone. The performance results show that with proposed approach, the coding gain can achieve the theoretical bound in some scenarios.**

I. Introduction

Network coding has recently emerged as a new paradigm that has demonstrated a practical way for improving the capacity of a network. In the literature, a number of network coding schemes have been proposed, including linear coding [1] and randomized coding [2]. Compared to these schemes, XOR-based coding (COPE) [3] is a simpler but effective way, which can enhance the capacity of multiple unicast traffic flows in wireless networks.

In a chain topology, assume nodes A and C need to exchange two packets P_1 and P_2. Due to communication range limitations, the intermediate node B serves as a relay node to forward the packets to their destinations. Because of radio interference, with traditional relay schemes, this information exchange takes four transmission slots to complete. The following is a possible sequence: 1) $A \mapsto B : P_1$; 2) $C \mapsto B : P_2$; 3) $B \mapsto C : P_1$; 4) $B \mapsto A : P_2$. In comparison, with a simple XOR-based network coding (as employed in COPE), this information exchange can be implemented by using three slots as in the following: 1) $A \mapsto B : P_1$; 2) $C \mapsto B : P_2$; 3) $B \mapsto A, C : P_1 \oplus P_2$. Because node A already has packet P_1, A can decode P_2 through the operation $P_2 = P_1 \oplus (P_1 \oplus P_2)$. Similarly, node C can decode P_1 through $P_1 = P_2 \oplus (P_1 \oplus P_2)$. With this network coding scheme, the number of transmission slots is reduced from 4 to 3, thus producing a coding gain of $\frac{4}{3} \cong 1.33$.



(a) Shortest path routing (b) Coding with backbone routing

Fig. 1. Example of network coding over a multi-hop wireless network

Despite the promising potential of network coding, a number of challenges have yet to be addressed. One of the most important issues is that current packet routing schemes in wireless ad hoc networks do not take advantage of network coding. Note that in COPE-type network coding schemes, when a transmission opportunity occurs at an intermediate node, the relay node would check the coding opportunity to combine multiple packets in the output queue together to broadcast with a single transmission. Because the route of the packets has multiple choices and the traffic flows are dynamic, the coding opportunity could be less with traditional ad hoc routing schemes. In Fig. 1, we illustrate the drawback of route selection with traditional routing schemes by a simple example. In this example, there are two flows in a multi-hop wireless network, one from node A to node D and the other from node E to node F. If the shortest path routing strategy is employed, the paths for the two flows are shown in Fig. 1(a). We can see that there is no coding opportunity if the routes shown in Fig. 1(a) are used.

To improve the efficiency of network coding, in this paper, we propose to apply network coding on a wireless backbone. In the literature, backbone construction has been investigated extensively to enhance the efficiency of the multi-hop wireless networks [4] [5]. To build a wireless backbone, some nodes with superior communication range, computational capability, or greater energy resources are selected as the *backbone nodes* (BNs). The main purpose of the BNs is to provide a wireless infrastructure facilitating network-wide efficient and resilient communications. Usually, a backbone construction algorithm utilizes a "connected dominating set" (CDS) of nodes to find a tree over the network graph, and a node not in the backbone has at least one backbone node neighbor (i.e., the BNs form a dominating set). Once a backbone has been constructed, any non-backbone node can access the nearest BN and achieve end-to-end communications through the backbone structure. Because the constructed backbone has a tree topology over the

[1]Any mention of commercial products in this paper is for information only; it does not imply recommendation or endorsement by NIST

Guo, Hui; Qian, Yi; Lu, Kejie; Moayeri, Nader.
"The Benefits of Network Coding over a Wireless Backbone."
Paper presented at IEEE Global Communications Conference (GLOBECOM 2009), Honolulu, HI, United States. November 30, 2009 - December 4, 2009.

network, the route of each packet can be simply determined at the source node.

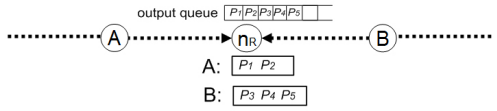
Our study is motivated by the observation that additional coding gain can be obtained if the packet forwarding is done in a way to maximize the coding opportunities at each hop of the backbone structure. In general, a backbone construction scheme will produce a connected backbone with a tree topology over the network, which provides a unique path for each pair of nodes. Hence, there will usually be a number of two-way flows over the backbone. Thus the coding opportunity increases dramatically because of the bi-directional nature of traffic. As shown in Fig. 1(b), if we construct a wireless backbone where the nodes F, G, H, C, D are selected as backbone nodes, then the two flows can "overlap" at the common nodes G, H and C, and then network coding can occur at these nodes. If the source nodes A and E continuously send packets to their destinations D and F respectively, network coding can be done continuously at the common nodes G, H and C. In addition, once a backbone is constructed, it is a semi-permanent wireless highway that can be established and maintained for some period of time, which reduces the computational cost of coding decisions.

In this paper, we explore the benefits of network coding over wireless backbones. The major contributions of this paper are: 1) To the best of our knowledge, this is the first work to investigate the gain of network coding over a wireless backbone; 2) We prove that the wireless backbone routing structure exhibits unique advantages for network coding, including increased coding opportunity and reduced overhead; 3) We conduct theoretical analysis of the proposed scheme and present numerical results; 4) To fully utilize the coding gain and avoid radio interference, we propose a simple optimized link scheduling algorithm and investigate its impact on the performance of wireless ad hoc networks.

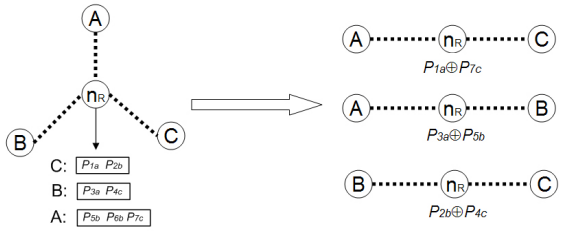
The rest of this paper is organized as follows. In Section II, we give an overview of the related work. In Section III, we present and analyze our network coding scheme over a wireless backbone. In Section IV, we propose an optimized link scheduling algorithm for a wireless backbone and analyze its performance. Finally, we conclude the paper in Section V.

II. Related Work

The use of network coding to enhance the capacity of a network was first proposed in [6] in the context of multicast communications. Subsequently, linear network coding [1] was shown to be sufficient for achieving optimal capacity based on the max-flow min-cut theorem. Since then the work [7] gave an algebraic characterization of linear encoding schemes, and distributed random network coding schemes were investigated in [2]. In the context of wireless networks, [8] studied the problem of minimum cost multicast involving a single session with a single source node. These works are all focused on multicast traffic. COPE is a pioneering work that presented a practical network coding scheme for multiple unicast traffic flows over multi-hop wireless networks [3].



(a) XOR coding for chain topology: n_R broadcasts $P_1 \oplus P_3$, $P_2 \oplus P_4$. The packet P_5 is left for the next round to be sent



(b) XOR coding for 3-star topology

Fig. 2. Examples of coding strategy for backbone nodes

However, the capacity improvement of COPE depends on the existence of coding opportunities, which in turn depends on the traffic patterns. To improve the coding opportunities in COPE-type coding approaches, several coding-aware routing schemes are proposed. Coding-aware routing is a routing paradigm that takes coding opportunities into account when performing route selection in network nodes. [9] and [10] proposed linear-programming-based coding-aware routing protocols. Although these routing protocols can improve the resource utilization and throughput, their centralized characteristic causes scalability problems. Therefore they are not practical to be employed in multi-hop wireless networks. In [11] and [12], the authors proposed distributed hop-by-hop coding-aware routing protocols, in which routing decisions are made based only on limited local view information. However, because of the dynamic routing strategy, when each node sends a packet, the node will locally calculate the coding opportunities to decide the next hop of that packet. Because packet transmission is the most frequent event, the computation at each node at high frequency is costly when conducting nexthop selection, which is especially true in dense wireless networks and heavy traffic scenarios. In addition, some of these schemes even require two-hop local view for information exchange, which results in even higher communication overhead.

Our proposed scheme for network coding over a wireless backbone can effectively overcome this issue. That is, it increases coding opportunity while not incurring any additional computational cost and communication overhead. To the best of our knowledge, this work is the first to propose using opportunistic network coding over a wireless backbone structure and to investigate the benefits of the scheme.

III. Network Coding over A Wireless Backbone

A. Coding Strategy

In earier work [5], we proposed a backbone construction scheme for heterogeneous wireless ad hoc networks. In our scheme, each BN has no more than three BN neighbors within its communication range, i.e., for each BN serving as a relay, it is part of either a chain topology or a 3-star topology, as

Guo, Hui; Qian, Yi; Lu, Kejie; Moayeri, Nader.
"The Benefits of Network Coding over a Wireless Backbone."
Paper presented at IEEE Global Communications Conference (GLOBECOM 2009), Honolulu, HI, United States. November 30, 2009 - December 4, 2009.

shown in Fig. 2. In our scheme, we have assumed that the multi-channel technique [13] is employed at the BNs, and thus there is no interference between BNs and regular nodes in network. As a result, for the channel used for backbone communications, there are at most three neighbors for each BN. Therefore, the capacity of the backbone can be improved compared with traditional network routing schemes with a single-channel system.

Fig. 2 shows two examples of *coding decision* (i.e., deciding which packets are encoded together to broadcast) in a chain topology (two BN neighbors) and a 3-star topology (three BN neighbors) respectively. In Fig. 2(a), there are 5 packets P_1 through P_5 buffered in the output queue of a relay node n_R, where the next hop of P_1, P_2 is node A and the next hop of P_3, P_4, P_5 is node B. Because the relay packets P_1 through P_5 either come from the node to *left side* of n_R or the node to *right side* of n_R, the previous hop of P_1, P_2 must be the node B and the previous hop of P_3, P_4, P_5 must be the node A. Also, it is certain that a packet P_i must exist in a neighbor of n_i if the neighbor is the previous hop of that packet. Therefore, n_R can broadcast the packets $P_1 \oplus P_3$ and $P_2 \oplus P_4$ in two transmission slots, and each packet is decodable at its intended nexthop. Because there is high likelihood of coding opportunities in bi-directional traffic, the remaining packet P_5 is left for the next round to be encoded with another packet.

A 3-star topology (which is possible at the backbone structure) can be decomposed into three chain topologies, as shown in Fig. 2(b). This example has seven packets queued at node n_R, where the next hop of P_1, P_2 is node C, the next hop of P_3, P_4 is node B and the next hop of P_5, P_6, P_7 is node A. In addition, the symbol P_{1a} means the previous hop of P_1 is the node A, and so forth. For each decomposed chain, n_R just combines two packets moving in opposite directions, thus the encoded packet can be decoded at their intended nexthops. Note that the remaining packet P_{6b} is left for the next round to be encoded with another packet.

We can see that this network coding scheme do not need the *reception reports* (which is required by the COPE) to let a node know what packets the neighbors have, because this information can be extracted by checking the previous-hop field in the packet header. Therefore, it dramatically reduces the overhead for information synchronization. In addition, the route between any source-destination pair is unique and predetermined with a virtual-backbone routing approach. Effectively, this results in a chain topology over which bidirectional traffic passes through. This increases coding opportunities. In summary, our proposed network coding over a wireless backbone scheme increases the coding opportunity while reducing the computational cost and overhead for coding decisions.

B. Coding Opportunity

In this subsection, we present a theoretical analysis of coding opportunity for both COPE and our scheme. Assume a given network node n_R has $N \geq 2$ neighbors within its communication range, and q packets buffered in its output queue. Using COPE-type opportunistic coding, the node n_R may find one or more packets in $\{P_2, P_3, ..., P_q\}$ to encode with the head of queue, P_1, and then broadcast them collectively, with the guarantee that the combined packets are decodable at their intended nexthops. We define the coding opportunity as the following:

Definition 3.1: If there exists at least one packet in the output queue that can be encoded with P_1, the head of output queue, and the encoded packet can be decoded to the native packets at corresponding intended nexthops, then there is a coding opportunity at node n_R.

We denote E_r as the event that there is a coding opportunity at node n_R. Because our scheme of network coding over a backbone does not need any reception reports to know what packets its neighbors have, for fair comparison (with the same computational cost and communication overhead), we assume n_R does not have access to reception reports and thus only two flows with reverse directions can be encoded together for transmission. In our analysis, we consider the scenario of uniform unicast traffic distribution, and we assume all nodes are independently and randomly deployed. We denote the previous hop of P_1 as N_i and the next hop of P_1 as N_j, and denote C_m the event that packet P_m has the reverse direction of P_1, i.e., $prev\_hop(P_m) = N_j$, $next\_hop(P_m) = N_i$. In addition, we define $P(C_m|n = N)$ as the probability of event C_m given that there is N number of neighbors around a node n_R, and $P(E_r|n = N, q)$ as the probability of event E_r given that there are N number of neighbors around a node n_R and the length of its output queue is q. Because of the uniform traffic distribution among the nodes, we have $P(C_m|n = N) = \frac{1}{N(N-1)}$, and the probability of its complement is given by $P(\overline{C_m}|n = N) = 1 - \frac{1}{N(N-1)}$. Hence, for an output queue of length q, we have $P(\overline{E_r}|n = N, q) = [1 - \frac{1}{N(N-1)}]^{q-1}$, and then

$$P(E_r|n = N, q) = 1 - [1 - \frac{1}{N(N-1)}]^{q-1} \quad (1)$$

Note that with the backbone structure, the number of BN neighbors is no more than three, i.e., $N \leq 3$.

Since the number of neighbors, N, is related to node density in a network without a backbone structure, next we investigate the coding opportunity with different node densities. Let ρ be the node density of the network, and R be the communication range of node n_R. We assume the nodes are deployed randomly according to Poisson point distribution, thus the probability of having n neighbors around n_R within a finite area (πR^2) is give by

$$P(n = N) = \frac{e^{-\rho\pi R^2}(\rho\pi R^2)^{(N+1)}}{(N+1)!} \quad (2)$$

then we have the probability of coding opportunity as a function of ρ and R

$$\begin{aligned} P(E_r|n \geq 2, q) &= \sum_{N=2}^{\infty} P(E_r, n = N|n \geq 2, q) \\ &= \frac{2\sum_{N=2}^{\infty}(1 - (1 - \frac{1}{N(N-1)})^{q-1})\frac{e^{-\rho\pi R^2}(\rho\pi R^2)^{(N+1)}}{(N+1)!}}{2 - 2e^{-\rho\pi R^2}\rho\pi R^2(1 + \rho\pi R^2)} \end{aligned} \quad (3)$$

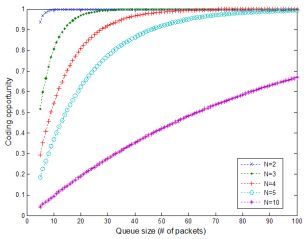


Fig. 3. Probability of coding opportunity as a function of queue size

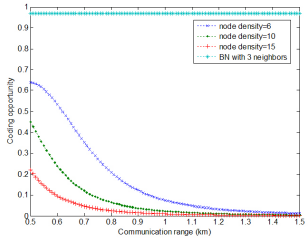


Fig. 4. Probability of coding opportunity as a function of communication range when $q = 20$

C. Numerical Results

Some numerical results are presented in Fig. 3, which is a plot of Eq. (1). We can see that the probability of coding opportunity increases with the output queue size. This indicates that there is a higher chance of coding opportunity in a network with heavy traffic. Thus network coding has more advantageous in heavy traffic scenarios, which is consistent with the intuition that network coding is not necessary in light traffic situations. Also, we can see that the probability of coding opportunity decreases with the number of neighbors. This demonstrates that chain topology (two neighbors) and 3-star topology (three neighbors) are more likely to lead to network coding opportunities, which is the case with our backbone topology presented in [5].

Fig. 4 shows the numerical results of Eq. (3) with $q = 20$. The results demonstrate that the probability of coding opportunity decreases with communication range. This is due to the fact that the number of neighbors increases with communication range. For fixed traffic intensity, the increase in number of nodes within communication range would decrease the coding opportunity. Whereas, for a backbone structure, the number of BN neighbors is not affected by the communication range. That is, the selected BNs within the communication range are always no more than four, because the number of BN neighbors is no more than three. Thus the probability of coding opportunity remains the same as communication range increases. In addition, we can see that for a network with no backbone structure, the probability of coding opportunity decreases with the node density.

IV. Link Scheduling and Observed Throughput

In this section, we investigate the scheduling issues of the backbone links and then derive an expression for *observed throughput* for relay nodes with an optimized link scheduling protocol. Here the term *observed throughput* means the amount of traffic passing through a specific relay node in a unit of time. In real-world, the theoretical coding gain 1.33 can only be achieved with optimal link scheduling. If the link scheduling is such that the transmitters always rotate as the following cycle: $A, C, B, ...$ (or $C, A, B, ...$), then node B can always encode two packets in each transmission and maximize the total throughput. However, if the link scheduling is $A, B, C, B, A, B, ...$, then node B cannot encode any packets. In practical situations, most of the wireless link scheduling

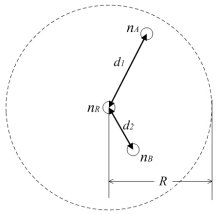


Fig. 5. An example for a BN (n_R) with two BN neighbors

if *the relay buffer of* n_R *is empty* **then**
  n_R pulls n_A to send a new packet $P_{(A->B)}$ to itself until the transmission is received successfully;
end
if *the relay buffer of* n_R *has a packet* $P_{(A->B)}$ *to be relayed to* n_B*, but has no packet* $P_{(B->A)}$ **then**
  n_R pulls n_B to send a new packet $P_{(B->A)}$ to itself until the transmission is received successfully;
end
if *the relay buffer of* n_R *has a packet* $P_{(B->A)}$ *to be relayed to* n_A*, but has no packet* $P_{(A->B)}$ **then**
  n_R pulls n_A to send a new packet $P_{(A->B)}$ to itself until the transmission is received successfully;
end
if *the relay buffer of* n_R *has both the packets* $P_{(A->B)}$ *and* $P_{(B->A)}$ **then**
  n_R broadcasts $P_{(A->B)} \oplus P_{(B->A)}$, which leads with the following probabilities to the situations listed:
  1) $(1 - p_{eA})(1 - p_{eB})$: both n_A and n_B receive the broadcast packet successfully, after which the relay buffer of n_R is empty;
  2) $p_{eB}(1 - p_{eA})$: n_A receives the broadcast packet successfully, but n_B does not, after which the relay buffer of n_R keeps $P_{(A->B)}$ for the next transmission;
  3) $p_{eA}(1 - p_{eB})$: n_B receives the broadcast packet successfully, but n_A does not, after which the relay buffer of n_R keeps $P_{(B->A)}$ for the next transmission;
  4) $p_{eA}p_{eB}$: neither n_A nor n_B receives the broadcast packet successfully, after which n_R rebroadcasts the combined packets in the next time slot.
end

Algorithm 1: A network coding aware link scheduling

protocols (such as a contention-based or random access link scheduling mechanism) are probabilistic in nature and not coding-oriented. Here we present a simple equal access protocol for backbone nodes in a chain topology that outperforms other sophisticated scheduling schemes by fully utilizing the potential coding opportunities. We then investigate the impact of lossy characteristics of a wireless channel on observed throughput of a relay node based on the proposed scheduling scheme.

Consider the general case where a BN (n_R) has two BN neighbors (n_A and n_B) and each one has packets that need to be relayed through n_R (i.e., n_R is an intermediate node to relay packets from n_A or n_B). The distance between n_A and n_R is d_1, and the distance between n_B and n_R is d_2, as shown in Fig. 5. This example also applies to the 3-star topology since a 3-star can be decomposed into three chain topologies, and the packets for exchanging between the two neighbors in a chain are a subset of the packets in the output queue based

Guo, Hui; Qian, Yi; Lu, Kejie; Moayeri, Nader.
"The Benefits of Network Coding over a Wireless Backbone."
Paper presented at IEEE Global Communications Conference (GLOBECOM 2009), Honolulu, HI, United States. November 30, 2009 - December 4, 2009.

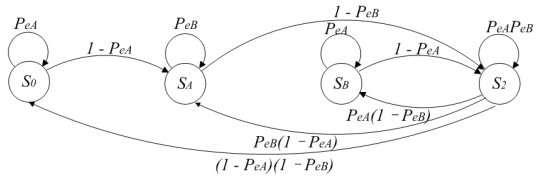


Fig. 6. Markov chain for state transitions of n_R

on their intended nexthops. Because the neighbors n_A, n_B usually have different distances from n_R ($d_1 \neq d_2$), when n_R broadcasts an encoded packet, n_A and n_B will receive that packet with different packet error rates (PER). Here we define the following terminology. p_{eA}: *PER of the link between n_A and n_R*; p_{eB}: *PER of the link between n_B and n_R*; pb_{eA}: *bit error rate (BER) of the link between n_A and n_R*; pb_{eB}: *bit error rate (BER) of the link between n_B and n_R*.

It is desirable that the relay node n_R always send XOR-ed packets in each transmission slot to fully utilize the broadcast nature of wireless medium. Here we present a simple pull-based link scheduling scheme in which n_R acts as a coordinator and pulls data from selected neighbors. This protocol is assumed to be used in heavy traffic scenarios, where the neighbor nodes n_A, n_B always have generated packets that need to be relayed by n_R. The protocol is described in Algorithm 1. The term *relay buffer* means a buffer space especially reserved for relay packets (i.e., n_R is neither the source nor the destination of the packets), the symbol $P_{(A->B)}$ denotes a packet with the direction from n_A to n_B. In this scheduling scheme, n_R coordinates its neighbors to satisfy the requirements of XOR-coding at each time slot. In this protocol, node n_R has four possible states, S_0: the relay buffer of n_R is empty; S_A: a packet $P_{A->B}$ is in n_R; S_B: a packet $P_{B->A}$ is in n_R; S_2: both packets $P_{A->B}$ and $P_{B->A}$ are in n_R. We can analyze the state transitions of n_R through the Markov chain shown in Fig. 6. Let $P(S_0)$, $P(S_A)$, $P(S_B)$ and $P(S_2)$ be the steady-state probabilities that n_R is in the states S_0, S_A, S_B and S_2 at a given time slot, respectively. Then we have

$$\begin{cases} P(S_0) = p_{eA}P(S_0) + (1-p_{eA})(1-p_{eB})P(S_2) \\ P(S_A) = p_{eB}P(S_A) + (1-p_{eA})P(S_0) \\ \qquad\quad +p_{eB}(1-p_{eA})P(S_2) \\ P(S_B) = p_{eA}P(S_B) + p_{eA}(1-p_{eB})P(S_2) \\ P(S_0) + P(S_A) + P(S_B) + P(S_2) = 1 \end{cases} \tag{4}$$

In each time slot, only when n_R is in the state S_2, it contributes to the observed throughput. Therefore the observed throughput is proportional to $P(S_2)$. From Eq. (4) we have

$$P(S_2) = \frac{1}{3 + (P_{eA} - P_{eB})^2/(1-P_{eA})(1-P_{eB})} \tag{5}$$

Here we assume that the duration of a time slot is T_0 seconds, and n_R can broadcast a packet of size L bits in a single transmission; then we define R_0 as $R_0 = \frac{L}{T_0}$. In a given time slot, if n_R is in state S_2 and the encoded packet $P_{A->B} \oplus P_{B->A}$ has been successfully received by both n_A and n_B, the achieved throughput in this time slot is $\frac{2L}{T_0} = 2R_0$. If either n_A or n_B receives the encoded packet successfully, the throughput in this slot is R_0. Therefore, we obtain the overall throughput with network coding as:

$$\begin{aligned} R_{nc} &= P(S_2)(p_{eA}(1-p_{eB}) + p_{eB}(1-p_{eA}) \\ &\quad + 2(1-p_{eA})(1-p_{eB}))R_0 \\ &= \frac{(2-p_{eA}-p_{eB})R_0}{3 + (P_{eA}-P_{eB})^2/(1-P_{eA})(1-P_{eB})} \end{aligned} \tag{6}$$

We then compute the throughput with a traditional transmission scheme for the example of Fig. 5, i.e., the intermediate node n_R just performs relay-and-forward for passing through packets. With a traditional transmission scheme, the node n_R will pull one of its neighbors (such as n_A) to send (or resend) the packet $P_{A->B}$ until n_R receives it correctly. Then n_R sends (or resends) the packet $P_{A->B}$ to n_B until n_B receives it successfully. The transmission from n_A to n_B takes $\frac{T_0}{1-p_{eA}} + \frac{T_0}{1-p_{eB}}$ seconds on average. This result is also applicable to the $P_{B->A}$ packet in the reverse direction of the same link. Thus, the observed throughput with a traditional transmission scheme is given by

$$R_{tr} = L/[\frac{T_0}{1-p_{eA}} + \frac{T_0}{1-p_{eB}}] = \frac{R_0(1-p_{eA})(1-p_{eB})}{(2-p_{eA}-p_{eB})} \tag{7}$$

and then the coding gain can be calculated by

$$C_{gain} = \frac{R_{nc}}{R_{tr}} = \frac{(2-p_{eA}-p_{eB})^2}{3(1-p_{eA})(1-p_{eB}) + (P_{eA}-P_{eB})^2} \tag{8}$$

We can see that when $p_{eA} = p_{eB} = p_e$, $R_{nc} = \frac{2}{3}(1-p_e)R_0$ and $R_{tr} = \frac{(1-p_e)R_0}{2}$, which confirms the previous analysis that the coding gain (R_{nc}/R_{tr}) is 1.33. In addition, we conclude that C_{gain} is always larger than 1 because $(2-p_{eA}-p_{eB})^2 - 3(1-p_{eA})(1-p_{eB}) - (p_{eA}-p_{eB})^2 = (1-p_{eA})(1-p_{eB})$, which is always larger than 0.

In this paper we model the lossy characteristics of wireless channel by the results in [14], where the authors derived the successful reception probability (SRP) as a function of distance (x) between a transmitter and a receiver under the log normal shadow fading model. The successful reception probability $SRP(x)$ can be approximated by

$$SRP(x) = \begin{cases} 1 - ((\frac{x}{K})^{2\beta})/2, & \text{if } x \leq K \\ ((\frac{2K-x}{K})^{2\beta})/2, & \text{if } K < x \leq 2K \\ 0, & \text{if } x > 2K \end{cases} \tag{9}$$

where K is the distance such that $SRP(K) = 0.5$, and β is the power attenuation factor ranging from 2 to 6. We define the communication range R as the following: within distance R from n_R, the $SRP(d)$ is larger than a threshold p_0. Because $SRP(R)$ (i.e., p_0) should be larger than 0.5 (i.e., $R < K$), for a fixed R, we have

$$K = \frac{R}{\sqrt[2\beta]{2(1-p_0)}} \tag{10}$$

In our analysis we assume that there is no error correction coding scheme employed. That is, if one of the bits in a packet is transmitted in error, the whole transmission is regarded as a failure. For a packet with L bits, we have $p_{eA} = 1 - (1 -$

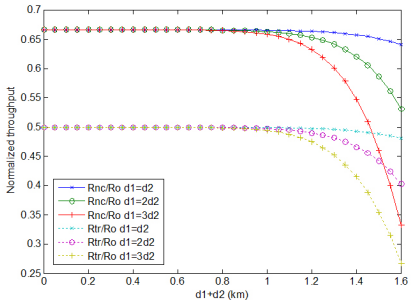


Fig. 7. Normalized observed throughput as a function of $d_1 + d_2$

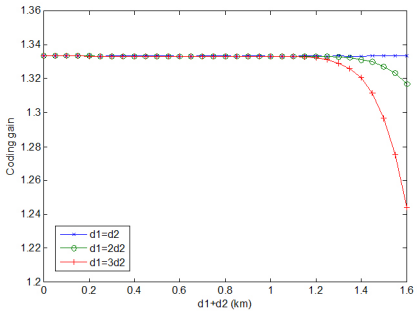


Fig. 8. Coding gain as a function of $d_1 + d_2$

$pb_{eA})^L$ and $p_{eB} = 1-(1-pb_{eB})^L$ where $pb_{eA} = ((\frac{d_1}{K})^{2\beta})/2$, and $pb_{eB} = ((\frac{d_2}{K})^{2\beta})/2$ based on Eq. (9).

Fig. 7 shows the results of the normalized observed throughput R_{nc}/R_0 and R_{tr}/R_0 as a function of $d_1 + d_2$. Here we assume the communication range (R) of n_R is 1.2 *Km*, $p_0 = 0.99$, L = 100 (*bits*) and $\beta = 4$. We also examine the impact of the asymmetry of the links $n_A \leftrightarrow n_R$ and $n_B \leftrightarrow n_R$ on the observed throughput by changing the ratio d_1/d_2. We can see that the observed throughput with network coding is always larger than that of a traditional transmission scheme (i.e., C_{gain} is always larger than 1), and when the two links are symmetric (i.e., $d_1 = d_2$), the system achieves the best observed throughput. In addition, the observed throughput is decreasing with the increasing of d_1+d_2, and with larger ratio d_1/d_2. In Fig. 8, we plot the results of Eq. (8) as a function of d_1+d_2. We can see that with asymmetric links (i.e., $d_1 \neq d_2$), the coding gain decreases with the increasing of d_1+d_2, which demonstrates that the asymmetry of lossy wireless links has significant impact on coding gain when d_1+d_2 is large enough. We can also see that in case of symmetric links, the coding gain always achieves the theoretical bound of 1.33.

V. Conclusions

In this paper, we have proposed a new paradigm for network coding over a wireless backbone and studied the benefits of the scheme. In our scheme, we construct a wireless backbone that facilitates end-to-end communications, and then employ an XOR-based network coding scheme over the backbone to further enhance the network efficiency and capacity. Because the wireless backbone provides bi-directional chain topology, which is particularly suitable for XOR-based network coding, the coding opportunity can be significantly improved compared with a COPE-type scheme over multi-hop wireless networks. In addition, our scheme does not introduce any additional computational cost and communication overhead compared with other coding-aware routing schemes. The theoretical analysis of coding opportunity in COPE is also presented and compared with our scheme. Furthermore, to fully utilize the coding gain, a simple optimized link scheduling protocol is presented and the observed throughput for a relay node is derived. The results demonstrate that with backbone routing and an optimized link scheduling scheme, network coding can effectively enhance the observed throughput for relaying nodes, and the coding gain achieves the theoretic bound 1.33 in symmetric ($d_1 = d_2$) or error-free links ($p_{eA} = p_{eB} = 0$). In addition, we find that the asymmetry of lossy wireless links has significant impact on network performance when the sum of d_1 and d_2 is large enough.

References

[1] S. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371-381, 2003.

[2] T. Ho, R. Koetter, M. Mdard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. of International Symposium on Information Theory*, Yokohama, Japan, Jun. 2003.

[3] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in The Air: Practical Wireless Network Coding," in *Proc. of ACM SIGCOMM*, 2006.

[4] A. Srinivas, G. Zussman, and E. Modiano, "Construction and Maintenance of Wireless Mobile Backbone Networks," *IEEE/ACM Transactions on Networking*, Vol.17, No.1, February 2009.

[5] H. Guo, Y. Qian, K. Lu, and N. Moayeri, "Backbone Construction for Heterogeneous Wireless Ad Hoc Networks," in *Proc. of IEEE ICC*, Dresden, Germany, June, 2009.

[6] R. Ahlswede, N. Cai, S. R. Li and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1024-1016, 2000.

[7] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782-796, October 2003.

[8] D. S. Lun, N. Ratnakar, M. Mdard, R. Koetter, D. R. Karger, T. Ho, and E. Ahmed, "Minimum-Cost Multicast over Coded Packet Networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2608-2623, June 2006.

[9] S. Sengupta, S. Rayanchu, and S. Banerjee, "An analysis of wireless network coding for unicast sessions: The case for coding-aware routing," in *Proc. of IEEE INFOCOM*, Anchorage, AK, USA, May 2007.

[10] B. Ni, N. Santhapuri, Z. Zhong, and S. Nelakuditi, "Routing with opportunistically coded exchanges in wireless mesh networks," in *Proc. of WiMesh06*, Reston, Virginia, USA, Sept. 2006.

[11] Y. Wu, S. M. Das, and R. Chandra, "Routing with a Markovian metric to promote local mixing," in *Proc. of IEEE INFOCOM*, Anchorage, AK, USA, May 2007.

[12] J. Le, J. C. S. Lui, and D. M. Chiu, "DCAR: Distributed coding-aware routing in wireless networks," in *Proc. of IEEE ICDCS*, Beijing, China, June 2008.

[13] A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. of IEEE INFOCOM*, Miami, FL, USA, May 2005.

[14] I. Stojmenovic, A. Nayak, J. Kuruvila, F. Ovalle-Martinez, E. Villanueva-Pena, "Physical Layer Impact on the Design and Performance of Routing and Broadcasting Protocols in Ad Hoc and Sensor Networks," *Computer Communications*, vol. 28, no. 10, pp. 1138-1151, June 2005.

Guo, Hui; Qian, Yi; Lu, Kejie; Moayeri, Nader.
"The Benefits of Network Coding over a Wireless Backbone."
Paper presented at IEEE Global Communications Conference (GLOBECOM 2009), Honolulu, HI, United States. November 30, 2009 - December 4, 2009.

POTENTIAL INTERFERENCE ISSUES BETWEEN FCC PART 15 COMPLIANT EMITTERS AND IMMUNITY COMPLIANT EQUIPMENT

Jeff R. Guerrieri, David R. Novotny, and Daniel G. Kuester
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80305

ABSTRACT

Transmitting equipment may interfere with sensitive electronic equipment even if both are in compliance with regulatory standards. This paper examines the potential for electromagnetic interference (EMI) between Federal Communications Commission (FCC) Part 15.247 for Ultra-High Frequency (UHF) emitters and immunity compliant sensitive equipment. At close ranges, the electromagnetic (EM) fields from these UHF emitters may exceed minimum standard immunity testing levels. This does not imply that interference will occur, but that the device may not be qualified to operate in the EM environment near the emitter. UHF Radio Frequency Identification (RFID) emitters are deployed regularly in locations where sensitive electronic equipment is in use. Recent studies have indicated that an interference potential can exist between some UHF emitters and medical, commercial and military systems. This paper estimates the range at which FCC compliant devices may pose a risk to industrial, consumer and medical devices and compares it to previously published data.

Keywords: FCC Part 15.247 devices, IEC 60601-1-2:2007 immunity levels, UHF RFID, 4W EIRP, sensitive electronic equipment

1.0 Introduction

This paper describes how interference could occur between an emitter operating within FCC Part 15 regulations and medical equipment that have passed required immunity tests. This could be a nuisance or have life threatening consequences.

Emitters of RF power must comply with regulations to mitigate interference with other devices. FCC Title 47 of the Code of Federal Regulations (CFR) Part 15 covers unlicensed radiators for Industrial, Scientific and Medical (ISM) equipment [1].

Transmit power limits are set in FCC Part 15 to balance between usable power for applications (telemetry, disturbance, communications, etc) and reasonable levels to minimize noise and interference potential. Section 15.247, which sets guidelines for frequency hopping and digitally modulated intentional radiators, allows up to 1 W conductor power to the antenna. Part 15 section 15.247 was drafted for UHF RFID [2] and other devices such as wireless smoke detectors and security systems that use frequency/power/modulation allocations [3].

Most electronic equipment must pass immunity requirements specific to its application. For example medical equipment and devices must meet the RF immunity requirements defined in IEC 60601 series and IEC 61000 series standards [4].

The electromagnetic fields that emanates from Part 15.247 UHF devices at close distances have the potential to be more intense than the minimum RF immunity levels that many consumer, industrial, and medical devices must operate satisfactorily. This does not imply interference will happen, only that the device may not be qualified to operate in the EM environment around these emitters. There are many different applications for RFID systems which give rise to the potential deployment in an uncontrolled environment. RFID systems have a greater potential of being used in very close range to a multitude of systems other than most other 15.247 emitters. Studies on medical devices have shown other UHF RFID systems may induce electric upset and the risks should be studied in greater depth [5,6].

This paper focuses on the potential effect of RFID emitters operating in the in the United States UHF ISM band, from 902 to 928 MHz, on electronic medical equipment.

The results will also be compared to results from a recent study in Europe [7]. The European Union (EU) UHF communications frequency band is from 864-868 MHz. Though the frequency allocations, power levels, channel and bandwidths differ slightly the basic analysis is applicable to the EU UHF devices. Similar EMI effects are being studied in other areas such as the military [8].

Guerrieri, Jeffrey; Novotny, David; Kuester, Daniel.
"Potential interference issues between FCC Part 15 compliant emitters and immunity compliant equipment."
Paper presented at 31st Antenna Measurement Techniques Association annual symposium 2009 (AMTA 2009), Salt Lake City, UT, United States. November 1, 2009 - November 6, 2009.

2.0 Unlicensed Radiator Limits

Unlicensed UHF, digital, frequency-hopping radios operate in the United States under 47 CFR Part 15 section 15.247 [9] and are restricted to 902-928 MHz. Section 15.247(b)(2) allows a "maximum conducted output power" or the maximum power to the antenna, P_T, of up to 1 W. At an output power of 1 W, the gain of the antenna, G_T, is limited to 6 dBi (decibel relative to isotropic). If the gain of the antenna is greater than 6 dBi, the power must be reduced "by the amount in dB that the directional gain of the antenna exceeds 6 dBi" **Error! Reference source not found.**. This sets the Effective Isotropic Radiated Power (*EIRP*) as defined below:

$$EIRP = P_T \cdot G_T \text{ (in Watts)}, \quad (1)$$

where the gain of the antenna is expressed in linear (non-dB) units,

$$G_T \text{ (linear units)} = 10^{\frac{G_T(\text{dBi})}{10}}. \quad (2)$$

The resulting maximum *EIRP* for UHF ISM is the often cited 4 W. This compares to EU limits of 2 W Effective Radiated Power (*ERP*). *ERP* and *EIRP* are related by a factor of 1.64 (the 2.15 dBi gain of a half-wavelength dipole):

$$EIRP = 1.64 \cdot ERP. \quad (3)$$

EU limits of 3.28 W EIRP are comparable to the US limits. The fields resulting from these radiators are discussed in Section IV of this paper.

3.0 Immunity Standards

Radiated RF immunity is a measure of the minimum RF field level at which a device must satisfactorily operate. Standards set immunity levels depending on product type, environment, and the requirements of the product to be in continuous operation. IEC 61000-6-1 sets immunity levels for general consumer devices, IEC 61000-4-3 sets immunity levels for general industrial devices and equipment, and IEC 60601-1-2 sets base radiated immunity for medical devices. Other product specific standards may set other radiated immunity standards as needed for specific classes of devices that have unique requirements [10]. These levels are summarized in Table 1.

Table 1. Minimum IEC RF Immunity Levels.

Standard	Radiated Immunity levels
IEC 61000-6-1:2005 Consumer Devices	3 V/m
IEC 61000-4-3:2008 Industrial Devices and Equipment	1 V/m – Level 1 Devices 3 V/m – Level 2 Devices 10 V/m – Level 3 Devices 30 V/m – Level 4 Devices
IEC 60601-1-2:2007 3rd Ed Medical Devices	3 V/m – Non Life-Critical Devices 10 V/m – Life-Critical Devices

Most critical, non-military devices must operate satisfactorily in the presence of a 3 or 10 V/m field. In reality, designers of critical products must make an assessment of worst-case levels and add an allowance for measurement uncertainty and test to that level [11]. The draft 4th Edition of IEC 60601-1-2 suggests that an actual analysis of the EM environment should be made and equipment designed and tested to that level. If the EM environment can't be assessed then the equipment should be tested to the 10 V/m level. Since this is still in draft most government agencies have not fully adopted the IEC 60601-1-2 4th Edition [11].

4.0 Radiated Field Levels

Using the Friis transmission equation, the far-field power density, P_D, in Watts/meter^2 from a radiator can be written as:

$$P_D = \frac{P_T G_T}{4\pi R^2} = \frac{EIRP}{4\pi R^2}, \quad (4)$$

where R is the distance from the transmitting antenna in meters. The far-field electric field amplitude, E, can be determined from the power density:

$$E = \sqrt{P_D \eta_0} = \sqrt{\frac{EIRP \cdot \eta_0}{4\pi R^2}} = \frac{\sqrt{EIRP \cdot 30}}{R}, \quad (5)$$

where η_0 is the impedance of free space. Equation 5 shows that the field halves every time distance is doubled. Near the antenna, less than a wavelength, the near-field effects may become significant and alter the field from Equation 5, and coupling between the antenna and the device may vary by other than $1/R$.

The calculated fields emanating from a 1 W source with a 6 dBi antenna (4 W EIRP) along with measured field levels are shown in Figure 1. The calculated and measured electric field levels are displayed as a function of distance from an RFID reader operating at 910 MHz.

Guerrieri, Jeffrey; Novotny, David; Kuester, Daniel.
"Potential interference issues between FCC Part 15 compliant emitters and immunity compliant equipment."
Paper presented at 31st Antenna Measurement Techniques Association annual symposium 2009 (AMTA 2009), Salt Lake City, UT, United States. November 1, 2009 - November 6, 2009.

These values are compared to the 3 and 10 V/m immunity test limits of 60601-1-2. Field levels greater than 10 V/m can be generated in the main beam of the emitter at distances less than 1 m. The less critical 3 V/m levels are experienced at 3.6 m from the antenna. This does not imply that the devices will be susceptible to interference closer than 1 or 3.6 m, respectively. It does show that the field levels illuminating a device may be higher than the immunity test levels.

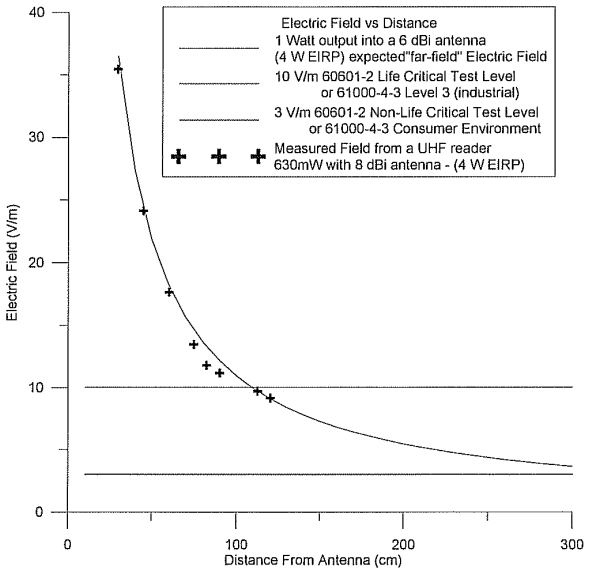


Figure 1. FCC Part 15.247 maximum allowable field level vs IEC 60601-1-2 minimum immunity field levels.

The predicted values in Figure 1 were confirmed using a commercial RFID reader set at 0.63 W conducted into the antenna port connected to a 8 dBi antenna (4 W EIRP). The RMS field level measured during RF carrier-on using a 15 cm field sensor is shown at various distances from the antenna taken inside a partially lined anechoic facility. There are slight variations between measured and calculated field levels. One possible explanation is that Equation 5 does not account for near-field effects and coupling between the source and probe.

5.0 Analysis of Interference

The gap between the emission and immunity standards revealed in the previous section could explain the inconsistent results in recent studies [5, 6, 7, 8]. Some testing and analysis in similar EU bands (865-868 MHz) have shown significant potential for interference [7, 8]. While testing of interference potential using some Section 15.247 equipment has shown only isolated effects on IEC 60601-1-2 compliant equipment [5, 6]. In both studies, the effects were most prominent in devices with voltage feedback or voltage detectors with moderate to high input impedances [5, 12]. Devices such as electrocardiogram (EKG) monitors could be more susceptible to stray signal interference as they have long high-impedance pickup devices and are measuring signals at the millivolt level.

Interference potential can also be affected by signal duty cycle, the rate of frequency change, and the range of the frequency hopping. FCC rules allows for essentially 100% carrier on time as long as all 50 channels are used equally. Duty cycle and rate of frequency hop vary between manufacturers and system usage. For example, some RFID devices interrogate with carrier-on duty cycles varying from under 25% to 100% depending on manufacturer, system complexity, and tag population. This variation in duty cycle can change the energy incident on a victim unit by a factor of 4:1. The interpretation of FCC frequency occupancy of 0.4 s for any 20 s period (consistent dwell for 0.4 s versus several shorter operations at a specific frequency) can have varying effects on frequency sensitive systems or the ability of fault detection routines to adapt. The smaller EU frequency band for UHF RFID and the listen before talk (LBT) frequency occupancy rule may also account for some of the differences between EU and US based immunity studies.

Modulated RF signals may cause interference at lower field levels if modulation correlates to the data rate of a victim device. Other tests have shown system vulnerabilities of equipment when systems are exposed to carriers modulated at frequencies near the data rate of the equipment [13]. In very rare cases, it is possible to interfere with equipment at much lower power levels than immunity standards dictate, especially if a modulation scheme affects a particular victim device. This can have the practical effect of increasing the range at which EM incidents occur. Standards have started to address the issues of modulated signal interference and some typical testing specifications to address this issue have been proposed [14]. However, they tend be limited to the frequency bands and modulation schemes of the telecommunications industry.

More study is needed on the immunity of critical systems, especially as the emissions allowance seems to conflict with immunity testing levels. IEC 61000-4-3 has testing requirements specific to cell phone and portable radio modulation types. The addition of a Part 15.247 emissions profile may be considered.

Guerrieri, Jeffrey; Novotny, David; Kuester, Daniel.
"Potential interference issues between FCC Part 15 compliant emitters and immunity compliant equipment."
Paper presented at 31st Antenna Measurement Techniques Association annual symposium 2009 (AMTA 2009), Salt Lake City, UT, United States. November 1, 2009 - November 6, 2009.

6.0 Recommendations

There are several ways to address the immunity versus emission issue. Detailed information of real interference potential is needed through EM environmental assessment and additional immunity studies. If the trends, as seen in [7], are borne out by additional studies, several possible actions could be considered. While the EMC community can wait to see if interference does arise in practical settings, other options can be taken to proactively address the potential immunity concerns.

Increase Immunity Requirements

60601-1-2:2007 already states the importance of immunity testing:

> *"[T]he existence of ELECTROMAGNETIC IMMUNITY standards is essential to assure safety of MEDICAL ELECTRICAL EQUIPMENT and MEDICAL ELECTRICAL SYSTEMS. ELECTROMAGNETIC COMPATIBILITY ... differs from other aspects of safety covered by IEC 60601-1 because the electromagnetic phenomena exist, with varying degrees of severity, in the normal use environment of all MEDICAL ELECTRICAL EQUIPMENT and MEDICAL ELECTRICAL SYSTEMS and by definition the equipment must "perform satisfactorily" within its intended environment in order to establish ELECTROMAGNETIC COMPATIBILITY. This means that the conventional single fault approach to safety is not appropriate for application to ELECTROMAGNETIC COMPATIBILITY standards [15]."*

After analysis of the EM environment is made for Section 15.427 devices and if a realistic concern is found, immunity standards for IEC 60601-1-2 and IEC 61000-4-3 might be upgraded to reflect the new conformance standards. It is reasonable to assume that the use of Section 15.247 compliant systems will increase (with some industry groups predicting the ubiquitous consumer portable RFID reader) but many other devices using the same allocation may become more prevalent.

Studies of early implantable devices highlighted immunity issues when exposed to UHF frequency RF emitters [16]. The realization of the existence of a UHF threat by telecommunications equipment may have resulted in implanted devices with improved signal line filtering. The EM hardening against UHF interference now used in many implantable devices may be part of the reason why recent studies showed much less vulnerability to UHF Section 15.247 based RFID compared to other conventional RFID frequencies [6].

In general, it is difficult to control the RF environment that any device experiences. Increased immunity standards in the IEC 60601-1-2: 2nd edition addressed this uncertainty and the 3rd edition of IEC 60601-1-2 goes further:

> *"This collateral standard also recognizes that for certain environments, higher IMMUNITY LEVELS may be required. Research necessary to determine how to identify the environments that may require higher IMMUNITY LEVELS, as well as what the levels should be, is in progress."[15]*

The negative aspect of changing the standard is that it takes time to write and accept standards. The US is currently using the 2nd Edition of IEC 60601-1-2 with the 2004 Amendment; the EU is using the 2nd Edition of IEC 60601-1-2 and will require compliance to the 2004 Amendment in 2009. Neither the US nor the EU has adopted the 3rd edition of 60601-1-2 (2007) [11]. Drafts for the 4th Edition of IEC 60601-1-2 are currently underway. In any new standard, the immunity of legacy devices needs to be considered.

Reduce UHF ISM 15.247 Power Allowances

UHF ISM designers and manufacturers, especially RFID system manufacturers, could work towards lower power requirements or regulation could be set to reduce radiated power levels. Researchers are proposing new designs for UHF RFID tags that require lower power [17, 18] that could be used at the same range with lower radiated power. Battery powered or battery-assisted tags may also reduce the required power levels – though tag lifetime becomes an issue. For a given antenna, reducing the power by a factor of four will reduce the electric field by a factor of 2 and the interference distance by a factor of 2, shown in Figure 2. It may not be practical to limit the power and range for an entire industry when some industry specific applications fall into this standards gap.

The prophesied ubiquitous reader in every store and in every cell phone may need to be examined more closely for allowed field levels in light of immunity concerns. However, it is very difficult to take back spectrum or power allowances once systems have been developed and deployed. Large investments have already been made by organizations with great lobbying power. Legacy devices will probably be exempt those systems will continue to use higher emitting devices.

Implement Usage Protocols for 15.247 Based Systems

A third option could be to implement regulations or "best practices". It may be reasonable to limit reader power in critical or susceptible environments. This would require documented administrative control or possibly engineering controls in the form of RF monitoring devices.

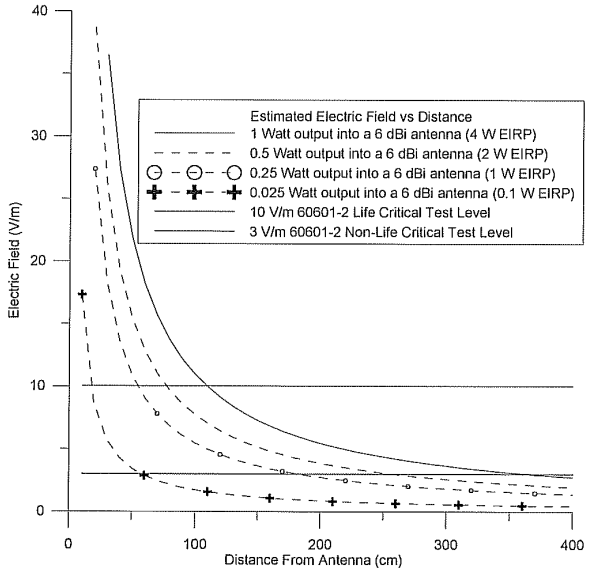


Figure 2. Estimation field levels vs. distance at different radiated power levels compared to IEC 60601-1-2 minimum immunity levels.

Operator Education

Operator implementations must to be considered. Many Section 15.247 authorized systems will allow for more than 1 W out of the source to account for losses in cable runs (allowable by FCC testing guidelines [19]). If used with lower loss or shorter cable, power at the antenna may exceed regulation. Manufacturer specified antennas may be replaced with higher gain models. This may also result in a higher radiated power. Combined, these effects can greatly increase EIRP, which could increase the interference potential, not to mention violating FCC regulations.

An unfamiliar operator may increase the output power of the reader if the desired communication performance is not met. Engineering controls limiting the power output of the reader could help mitigate this problem. Well documented usage and deployment guidelines along with regulation enforcement may help reduce interference potential. The responsibility lies with the employer and operator to ensure that the equipment is operating within limits set by regulation.

7.0 Conclusions

It is difficult to predict the interference potential of regulated emitters on immunity compliant devices.

FCC Part 15.247 regulates UHF emitters to operate at less than 4W EIRP. At distances of less than 1 m these regulated devices will emit field levels greater than 10 V/m.

The highest immunity levels set by IEC 60601-1-2:2007 for life critical devices are 10 V/m. This suggests that an interference potential exists within 1 m of Part 15.247 devices. The IEC 61000 series sets immunity levels for consumer and industrial devices. For Level 1, 2 and 3 devices the immunity field strengths are 1, 3, and 10 V/m, respectively. Only Level 4 devices are tested to 30 V/m. At distance less than 1 m to Part 15.247 emitters there exists a potential for interference.

It is recognized within IEC 60601-1-2 that these issues are the responsibility of the victim unit, the operator, and the interfering device.

While much of the current research is focused on RFID implementations [20], the same potential may exist for many other Part 15.247 devices. Wireless communication systems, wireless smoke detector systems and a multitude of other wireless systems have similar interference potential to immunity compliant devices. These other wireless systems have the potential to interfere with an RFID system also.

8.0 References

[1] FCC 47 CFR Part 15 – 10 July 2008.

[2] S. Dayhoff, "NEW POLICIES FOR PART 15 DEVICES," TCBC Workshop,13-15 May 2005.

[3] FCC rules ET Docket No 03-201 – Adopted 8 July 2004.

[4] IEC 61000-4-2 Edition 1.2 2001-04, IEC 60601-1-2: Second Edition 2001 , IEC 60601-1-2 Third Edition 2007-03 and IEC 61000-6-2: Second Edition 2005-01.

[5] B, Christe, E. Cooney, G. Maggioli, D. Doty, R. Frye, J. Short, "Testing Potential Interference with RFID Usage in the Patient Care Environment," Biomedical Instrumentation & Technology: Vol. 42, No. 6, pp. 479–484, Nov. 2008.

[6] S. Seidman, P. Ruggera, R Brockman, B Lewis, M Shein, "Electromagnetic compatibility of pacemakers and implantable cardiac defibrillators exposed to RFID readers," Int. J. Radio Frequency Identification Technology and Applications, Vol. 1, No. 3, pp. 237-46, 2007.

[7] Remko van der Togt; Erik Jan van Lieshout; Reinout Hensbroek; E. Beinat; J. M. Binnekade; P. J. M. Bakker, "Electromagnetic Interference From Radio Frequency Identification Inducing Potentially Hazardous Incidents in Critical Care Medical Equipment", Journ of the American Medical Association, June 25, 2008; vol. 299: 2884 – 2890.

[8] A. Moro, "Potential Interference Generated by UHF RFID systems on Military Telecommunication Devices," EU RFID FORUM 2007, March 13th and 14th 2007, Albert Borschette Conference Centre, Brussels http://www.rfidconvocation.eu.

[9] FCC 47 CFR Part 15 – 10 July 2008 – section 15.247(b)(4)

[10] ISO 60601-2-12:2001 "Medical electrical equipment -- Part 2-12: Particular requirements for the safety of lung ventilators -- Critical care ventilators" deals with special safety requirements for a specific class of medical equipment.

[11] K. Armstrong, "Medical EMC Standards: Towards the 4th Edition of IEC 60601-1-2" Conformity, Vol 13, Issue 9, pp 20-28, Sept 2008.

[12] Detail of data presented in [6] located at http://www.amc.nl/?pid=5266.

[13] D. Novotny, J. Guerrieri, M. Francis, K. Remley,"HF RFID Emissions and Electromagnetic Performance," IEEE International Symposium on Electromagnetic Compatibility, Detriot 2008.

[14] IEC 61000-4-3:2006+A1:2007 : Annex A "Rationale for the choice of modulation for tests related to the protection against RF emissions from digital radio telephones".

[15] IEC 60601-1-2:2007 3rd Edition, pp. 16-17.

[16] R.E. Schlegel, F.H. Grant, "Wireless phones and cardiac pacemakers: in vitro interaction study," Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE, 30 Oct.-2 Nov. 1997, vol.6, Page(s):2551 - 2554.

[17] A. Ricci, M. Grisanti, I. De Munari, P. Ciapolini, "Design of a Low-Power Digital Core for Passive UHF RFID Transponder," 9th EUROMICRO Conference on Digital System Design (DSD'06), 2006, pp.561-568,.

[18] U. Karthaus and M. Fischer, "Fully integrated passive UHF RFID transponder IC with 16.7-uW minimum RF input power," IEEE Journal of Solid-State Circuits, October 2003, 38(10):1602–1608.

[19] FCC Public Notice DA 00-705 "Filing and Measurement Guidelines for Frequency Hopping Spread Spectrum Systems", 30 March 2000.

[20] B.S. Ashar; A. Ferriter, "Radiofrequency Identification Technology in Health Care: Benefits and Potential Risks," Journ of the American Medical Association, November 21, 2007; 298: 2305 – 2307.

Guerrieri, Jeffrey; Novotny, David; Kuester, Daniel.
"Potential interference issues between FCC Part 15 compliant emitters and immunity compliant equipment."
Paper presented at 31st Antenna Measurement Techniques Association annual symposium 2009 (AMTA 2009), Salt Lake City, UT, United States. November 1, 2009 - November 6, 2009.

Traceability: The Key to Nanomanufacturing

N.G. Orji, R.G. Dixson, B.M. Barnes, and R. M. Silver
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899, USA

Abstract: *Over the last few years key advances have been made in the area of nanomanufacturing and nanofabrication. Several researchers have produced nanostructures using either top-down or bottom-up techniques, while other groups have functionalized such structures into working devices. In all cases, for the devices to be useful, there has to be a way not only to measure their properties but also to know that the results are valid. The measurements have to be traceable. Some of the properties of these nanostructures are determined by size; this increases the importance of accurate measurements. In this paper, we present work that we have done to introduce traceability to a nanomanufacturing environment, using a concept called reference measurement system. The paper focuses on length traceability.*

Keywords: Traceability, Nanomanufacturing, reference measurement system, critical dimension, Scatterfield microscopy

1. Introduction

As fundamental research has now yielded standalone nanostructures using top-down techniques[1], and bottom-up techniques such as self-assembly[2], nanomanufacturing and nanofabrication are emerging as key industries for the years ahead. The key objective is to make nanoscale devices that perform certain functions. One of the most important parameters of nanostructures is size. This is because functionality at the nanoscale is determined by size. To ensure that the dimensional measurements are valid, there has to be a way to validate the results. This is done through traceability to established standards. In areas such as microlithography, the features being manufactured are increasingly getting smaller. The international technology roadmap for semiconductors (ITRS) 2007 edition[3] forecasts that the half pitch of DRAM features will be around 25 nm by the year 2015. This puts a burden on the resolution of the instruments needed to measure such small features[4, 5]. In addition to instrument resolution, traceability is also important. This ensures that 25 nm is indeed 25 nm.

In this paper we will describe work we have done to introduce traceability to a semiconductor nanomanufacturing environment[6-8]. The focus is on length traceability. The paper is organized as follows: In section two we will give a brief overview of how traceability is realized for most length measurement instruments. This is followed by a description of a reference measurement system in section three. In section four we will outline how *SI* traceability is realized for one of the measurands characterized by the reference instrument, while section 6 describes current work in Scatterfield optical microscopy, and how AFM measurements are used to reduce the uncertainty. The paper will conclude with a discussion and summary.

2. Traceability in Length Metrology

The Vocabulary of Basic and General Terms in Metrology (VIM)[9] defines traceability as "the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties."

In length metrology, the primary reference is the *SI* (*Système International d'Unités* or International System of Units) unit of length - the meter. For displacement measurement instruments, one of the ways to achieve traceability is to monitor the motions of a scanning system within a defined coordinate system using displacement interferometry. Usually a laser with a 633 nm wavelength of the I_2-stabilized He-Ne laser (a recommended radiation for the realization of the meter in the visible wavelength) is used[10]. Other ways of introducing traceability include the use of atomic lattice spacing[11-13]. The lattice spacings are

Sixth International Conference on "**Precision, Meso, Micro and Nano Engineering**" 11-12 December 2009.

Orji, Ndubuisi; Dixson, Ronald; Barnes, Bryan; Silver, Richard.
"Traceability: The Key to Nanomanufacturing."
Paper presented at Sixth International Conference on Precision, Meso, Micro and Nano Engineering, Coimbatore, India. December 11, 2009 - December 12, 2009.

measured using x-ray diffraction[14], and can serve as length standards at the nanoscale. For nanoscale measurements, traceability is important when comparing the performance or consistency of results from different instruments, and when instruments based on different technologies are used. Traceability to a standard is also important when verifying the reliability of models or validating the results of work performed at a different location.

One of the industries (among others) where length traceability is important is the semiconductor industry. Traditionally, the semiconductor industry placed greater emphasis on precision rather than traceability. However, as feature sizes continue to decrease; the performance of devices depends not only on precision, but also on accuracy. Recent results linking parameters such as line edge roughness[15-18] to device performance underscore this[19].

3. Reference Measurement System

To address the issue of lack of traceability in semiconductor length measurements, the National Institute of Standards and Technology and SEMATECH (a semiconductor industry consortium) implemented a Reference Measurement System (RMS). The RMS is a well characterized and traceable instrument that is used to evaluate the performance of tools used in a semiconductor manufacturing facility[20-23]. The ITRS[3] describes an RMS as an instrument that "is well characterized using the best science and technology of dimensional metrology can offer: applied physics, sound statistics, and proper handling of all measurement error contributions."

Figure 1 shows a schematic diagram of the RMS concept. A RMS starts with calibration standards or first principles realization of the definition of the meter[7, 24, 25]. The standards are used to characterize one of the better performing instruments, which is then used to evaluate a set of reference wafers. The wafers are used to evaluate the performance of an in-line tool. Preferably, there will be reference wafers for each product the fabrication facility measures. This ensures that the measurements made using the RMS instrument are consistent with the final product. The RMS could be one or more instruments[26, 27]. A key requirement is that it should have better performance than the instruments it is evaluating.

The specific instrument described in the rest of the paper is a critical dimension atomic force microscope (CD-AFM)[20-22]. The CD-AFM is a specialized atomic force microscope that actuates and senses in two directions instead of one[28]. For dimensional metrology applications in the semiconductor fab, the CD-AFM works well as a reference instrument because it is relatively insensitive to the different materials being measured. The instrument is calibrated for height, pitch, linewidth, and sidewall angle[29]. These measurements lend *SI* traceability to a wide range of production relevant samples[6, 7, 12]. One of the key measurements in semiconductor dimensional metrology is linewidth (also known as critical dimension). It currently represents the smallest feature that needs to be controlled in the semiconductor lithography process. The rest of the paper shows how we use the CD-AFM to calibrate the linewidth standards[23, 25, 30], and reduce the uncertainty of other instruments.

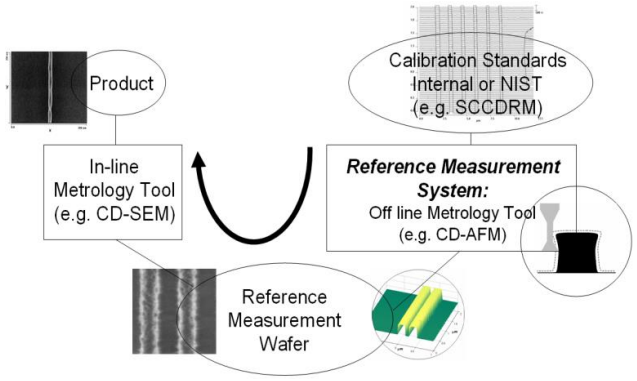


Figure1: Schematic diagram of the Reference Measurement System concept.

Orji, Ndubuisi; Dixson, Ronald; Barnes, Bryan; Silver, Richard.
"Traceability: The Key to Nanomanufacturing."
Paper presented at Sixth International Conference on Precision, Meso, Micro and Nano Engineering, Coimbatore, India. December 11, 2009 - December 12, 2009.

4. First Principles CD-AFM Tip Calibration for Linewidth

To measure linewidths with the CD-AFM, the size of the tip should be known. This is because the CD-AFM uses specialized tips known as "boot-shaped" or "flared" tips to access the sidewall[25, 31]. Figure 2(a) shows a schematic diagram of a CD-AFM tip accessing a feature sidewall. The protrusions at the lateral end of the tip are needed to make contact with the feature. By contrast, figure 2(b) shows a conical AFM tip measuring a patterned feature. The profile traced by the conical tip does not faithfully represent the measured feature. However, the profile traced by the boot-shaped tip adequately represents the features, but the width is larger by the size of the tip. Knowing the size of the tip *a priori* allows the user to determine the actual size of the feature. Figure 3 shows the tip width determination sequence. TW stands for Tip Width, CW stands for characterizer width, and AW represents the apparent width. The primes indicate known dimensions. The TW is unknown but the width of the tip calibration structure CW' is known. The apparent width AW' produced by CW' and TW is known. To get TW, CW' is subtracted from AW'. This relatively straight forward approach belies the substantial work required to calibrate CW. For measurements of linewidth, the process is reversed.

The most accurate method used to evaluate the characterizer width is lattice spacings imaged using transmission electron microscope. Descriptions of our previous work using high resolution transmission electron microscope (TEM) are well documented[32]. Here we describe supporting work we are doing with a different mode of TEM to validate the results of the initial measurements[13, 32].

Two modes of TEM (among others) that produce crystal lattice-traceable images are high resolution transmission electron microscope (HR-TEM) and high angle annular dark field scanning transmission electron microscope (HAADF-STEM). HR-TEM produces lattice-traceable images by interference patterns of the diffracted and transmitted beams rather than the actual atomic columns, while HAADF-STEM produces direct images of the crystal lattice[33]. To calibrate feature width, the number of lattice spacings from edge to edge has to be determined. One of the key uncertainty sources in this exercise is edge determination. For the purposes of dimensional calibration, the difference in how these two modes of TEM work could cause subtle variations in the way feature edges are defined. We wanted to quantify this variation. A HR-TEM image of a line feature is shown in figure 4, while figure 5 shows a HAADF-STEM image of the same feature. The roll-off at the left edge of the image in figure 5 could be due to aberration in the optics. The images were taken from the middle of the feature.

The cross-sectional size of the sample used was approximately 18 nm, and determined by the HAADF-STEM, which required a field of view of < 25 nm in order to see the atomic lattice. The sample was aligned for the z = [112] of the silicon substrate. The HAADF-STEM image was taken first followed by the HR-TEM. The field of view of the instrument did not allow the whole structure to be measured at high resolution. The spacing in undoped Si (111) is 0.313560137 nm with a standard uncertainty of ± 0.000 000 009 nm, as determined from X-ray diffraction[14].

The TEM images shown here are imaged as lines rather than individual atoms. This makes it easier to count the atomic columns. Also shown in figures 4 and 5 are profiles extracted from the TEM images. These profiles are used to determine the width. The highlighted portions of the profiles indicate the lattice positions that contribute to the uncertainty.

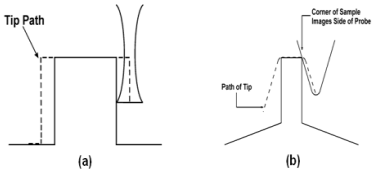


Figure 2: Schematic diagram of (a) CD-AFM tip scanning a feature, (b) conventional AFM tip scanning a feature.

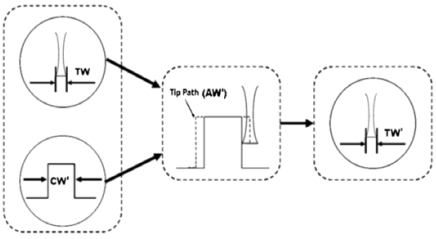


Figure 3: Schematic diagram of the tip-width determination sequence.

Orji, Ndubuisi; Dixson, Ronald; Barnes, Bryan; Silver, Richard.
"Traceability: The Key to Nanomanufacturing."
Paper presented at Sixth International Conference on Precision, Meso, Micro and Nano Engineering, Coimbatore, India. December 11, 2009 - December 12, 2009.

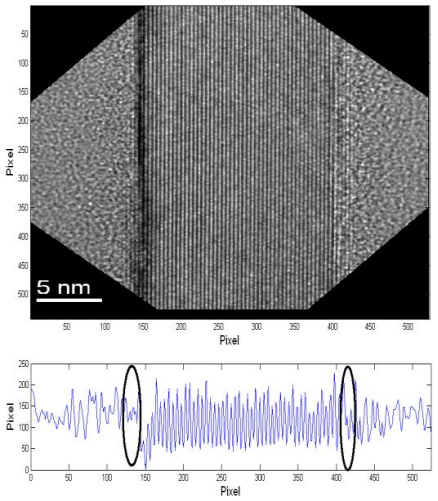


Figure 4: HR-TEM image of a SCCDRM feature. The profile is from the center of the HR-TEM image. The questionable edge locations are highlighted.

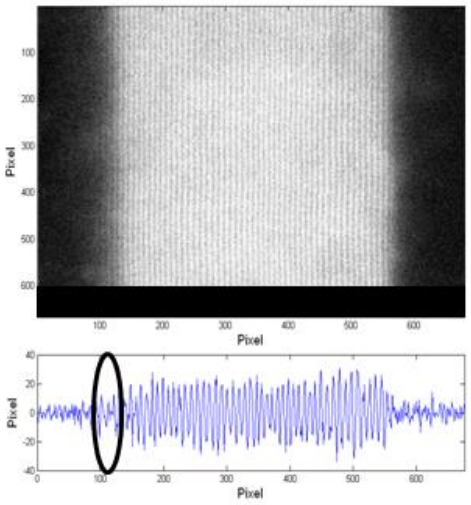


Figure 5: HAADF-STEM image of a SCCDRM feature. The profile is from the center of the HR-TEM image. The questionable edge locations are highlighted.

5. Scatterfield Optical Microscopy

One of the uses of the RMS is to calibrate other instruments used in the semiconductor manufacturing research. One such instrument is the Scatterfield microscope. There has been significant recent research investigating new optical technologies for critical dimension and overlay metrology for manufacturing at and beyond the 32 nm node. Much of this work has focused on scatterometry and, more recently, scatterfield microscopy, a technique combining well-defined angle-resolved illumination with image-forming optics. This has been summarized in Silver *et al.*[34] and in the references contained therein. These optical methods are of particular interest because of their nondestructive, high-throughput characteristics and their potential for excellent sensitivity and accuracy.

The scatterfield-microscopy instrument is based on a Köhler illuminated bright-field microscope, such that each point at the conjugate back focal plane maps nominally to a plane wave of illumination at the sample. Access to a large conjugate back focal plane enables engineered illumination that has resulted in further advances in optical-system characterization and data analysis. As a result, the microscope-illumination and collection-path errors have been mapped to a functional dependence. They can then be used to normalize the experimental data for accurate comparison with electromagnetic simulations. We acquire both microscope images and backgrounds as a function of angle and calculate the mean intensity of the angle-resolved images, which have been corrected using the background scan that was previously normalized by the known silicon reflectance. This is similar to conventional scatterometry, except that measurements were made with high-magnification image-forming optics.

We compared the normalized experimental signatures with electromagnetic scattering simulations using parametric analysis. We assembled a library of curves by simulating a multidimensional parameter space. We completed the comprehensive simulations using a rigorous coupled-waveguide analysis (RCWA) model. Figure 6 shows angle resolved experimental data and fitting results that demonstrate the accuracy that can now be achieved using this method. Good agreement between the simulated library of curves and the experimental data is observed and residuals are acceptable. However, the fitting process produces more uncertainty than desired, 1 to 2nm (1σ).

Measurement uncertainties for optical scatterometry and Scatterfield measurements are fundamentally limited by the underlying cross-correlations between the different fit parameters, *e.g.,* line widths and

Orji, Ndubuisi; Dixson, Ronald; Barnes, Bryan; Silver, Richard.
"Traceability: The Key to Nanomanufacturing."
Paper presented at Sixth International Conference on Precision, Meso, Micro and Nano Engineering, Coimbatore, India. December 11, 2009 - December 12, 2009.

heights. To reduce parametric correlation and improve measurement performance and uncertainties, we have developed a Bayesian statistical approach[35, 36] that integrates *a priori* information gleaned from other measurements. This allows us to embed information obtained from reference metrology and complimentary ellipsometry of the optical constants. The Bayesian embedded metrology approach was applied to the silicon target used in Figure 6. Table 1 shows the best-fit values and uncertainties for the regression analysis, with and without embedded atomic-force-microscopy (AFM) reference metrology. The data show a change in the mean values as well as an improvement in the uncertainties.

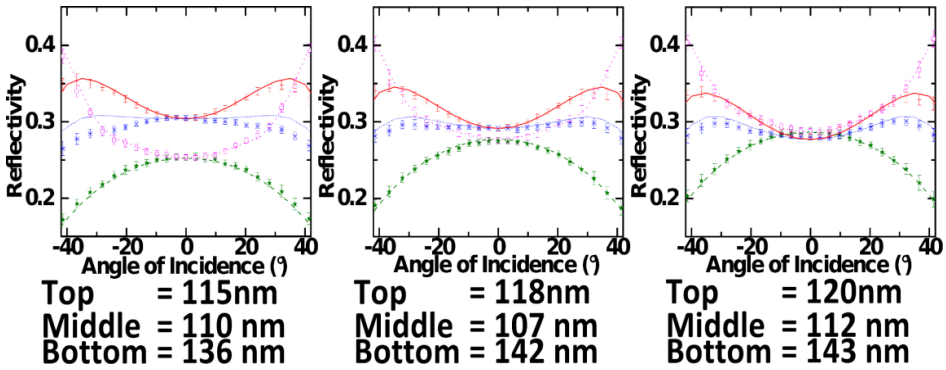


Figure 6: Experimental data and library data fits for three measurements. Top and middle critical dimensions (CDs) show good agreement with reference values.

Table 1: Optical CD (OCD) measurement results with and without embedded atomic-force microscopy (AFM) reference metrology.

	OCD Fitting	AFM	OCD with AFM
CD_{Top}	120	119.2	121
CD_{Middle}	112	117.3	115
CD_{Bottom}	143	132.8	141

	OCD Fitting	AFM	OCD with AFM
σ_{Top}	1.05	0.75	0.35
σ_{Middle}	1.58	0.75	0.60
σ_{Bottom}	0.78	0.75	0.42

6. Discussion

The above example highlights a way to introduce traceability through the Si lattice spacing. The number of lattice positions close to the edge that we could not conclusively resolve in figure 6 is three. The error associated with the edge definition can be substantially reduced by using the average of several measurements rather than one[37]. An uncertainty of three lattice positions translates into ≈ 0.94 nm for edge determination component. To put this in the larger context, the ITRS forecast a CD uncertainty of 0.28 nm (3σ) for the year 2010 and 0.22 nm (3σ) for 2012. This is the type of improvement needed to continue to make functional devices that adhere to Moore's law. Currently there are no known methods to meet these requirements. One of the approaches we are pursuing is the use of aberration corrected TEM. This will substantially reduce the uncertainty due to aberration of the electron optics. The uncertainty of determining the feature width for CD-AFM calibration is currently less than 1 nm[32, 38]. In addition to determining the sample width, other uncertainty sources include those associated with the CD-AFM instrument[20, 38].

The Scatterfield microscopy example shows how results from the reference instrument are used to reduce the uncertainty of other tools. In this case the Scatterfield microscope is faster, and better suited for inline use. Incorporating the AFM results has the benefit of decoupling terms that are correlated in the optical measurement, and thereby reducing the uncertainty[34].

Orji, Ndubuisi; Dixson, Ronald; Barnes, Bryan; Silver, Richard.
"Traceability: The Key to Nanomanufacturing."
Paper presented at Sixth International Conference on Precision, Meso, Micro and Nano Engineering, Coimbatore, India. December 11, 2009 - December 12, 2009.

The above examples highlight the type of calibration exercise needed to ensure traceability at the nanoscale. Overall the gap between what is required by the ITRS, and what is possible is large. Nanomanufacturing as a whole will benefit greatly from incorporating traceability. This will facilitate replication of results and identification of problems when they occur. Some of the strategies used to introduce *SI* traceability in semiconductor dimensional metrology are directly applicable to the burgeoning nanomanufacturing industry.

7. ACKNOWLEDGMENTS

The authors are grateful to R. Attota and Y. Obeng for their comments and suggestions. This work is partially supported by the Office of Microelectronics Programs (OMP) and the Nanomanufacturing Program at NIST.

REFERENCES

1. Austin, M.D., et al., Fabrication of 5 nm linewidth and 14 nm pitch features by nanoimprint lithography. *Applied Physics Letters*, **84**(26): p. 5299-5301, (2004).

2. Cui, Y. and Lieber, C.M., Functional nanoscale electronic devices assembled using silicon nanowire building blocks. *Science*, **291**(5505): p. 851-853, (2001).

3. *International Technology Roadmap for Semiconductors (ITRS), Metrology Chapter*, San Jose, (SIA, 2007).

4. Knight, S., et al., Advanced metrology needs for nanoelectronics lithography. *Comptes Rendus Physique*, **7**(8): p. 931-941, (2006).

5. Smith, S., et al., Comparison of measurement techniques for linewidth metrology on advanced photomasks. *IEEE Transactions on Semiconductor Manufacturing*, **22**(1): p. 72-79, (2009).

6. Orji, N.G., Dixson, R.G., Cordes, A., Bunday, B.D., and Allgair, J.A. Measurement traceability and quality assurance in a nanomanufacturing environment. in *Proc. SPIE* 7405, 740505 (SPIE, 2009). https://doi.org/10.1117/12.826606

7. Orji, N., Dixson, R., Bunday, B., and Allgair, J. Accuracy considerations for critical dimension semiconductor metrology. in *Proc. SPIE* 7042, 70420A (SPIE, 2008). https://doi.org/10.1117/12.796372

8. Dixson, R., et al. Traceable atomic force microscope dimensional metrology at NIST. in *Proc. SPIE* 6152, 61520P (SPIE, 2006). https://doi.org/10.1117/12.656803

9. BIPM, *International Vocabulary of Basic and General Terms in Metrology (VIM)*, Geneva, (BIPM, 2008).

10. Dixson, R., Koning, R., Fu, J., Vorburger, T., and Renegar, B. Accurate dimensional metrology with atomic force microscopy. *Proc. SPIE* 3998, (SPIE, 2000).https://doi.org/10.1117/12.386492

11. Dixson, R., et al. Silicon single atom steps as AFM height standards. in *Proc. SPIE* 4344, (SPIE, 2001). https://doi.org/10.1117/12.436739

12. Orji, N.G., Dixson, R.G., Fu, J., and Vorburger, T.V., Traceable pico-meter level step height metrology. *Wear*, **257**(12), (2004).

13. Orji, N.G., et al. TEM calibration methods for critical dimension standards. in *Proc. SPIE* 6518, 651810 (SPIE, 2007). https://doi.org/10.1117/12.713368

14. Mohr, P.J. and Taylor, B.N., CODATA recommended values of the fundamental physical constants: 1998. *Journal of Physical and Chemical Reference Data*, **28**(6): p. 1713-1852, (1999).

15. Bunday, B.D., et al. Determination of optimal parameters for CD-SEM measurement of line-edge roughness. in *Proc. SPIE* 5375, (SPIE, 2004). https://doi.org/10.1117/12.535926

16. Orji, N., Sanchez, M., Raja, J., and Vorburger, T., AFM characterization of semiconductor line edge roughness, in *Applied Scanning Probe Methods*. p. 277-301,(Springer Berlin Heidelberg, 2004).

17. Orji, N.G., et al., Line edge roughness metrology using atomic force microscopes. *Measurement Science and Technology*, **16**(11), (2005).

18. Vorburger, T., Fu, J., and Orji, N., *In the rough*, in *Opt. Eng. Mag*. p. 31-34, (SPIE, March, 2002).

19. Diaz, C.H., Tao, H.J., Ku, Y.C., Yen, A., and Young, K., An experimentally validated analytical model for gate line-edge roughness (LER) effects on technology scaling. *Ieee Electron Device Letters*, **22**(6): p. 287-289, (2001).

20. Dixson, R.G., Guerry, A., Bennett, M.H., Vorburger, T.V., and Bunday, B.D. Implementation of reference measurement system using CD-AFM. in *Proc. SPIE* 5038, (SPIE, 2003). https://doi.org/10.1117/12.483667

21. Dixson, R. and Guerry, A. Reference metrology using a next generation CD-AFM. in *Proc. SPIE* 5375, (SPIE, 2004). https://doi.org/10.1117/12.536898

22. Orji, N.G., Martinez, A., Dixson, R.G., and Allgair, J. Progress on implementation of a CD-AFM-based reference measurement system. in *Proc. SPIE* 6152, 61520O (SPIE, 2006). https://doi.org/10.1117/12.653287

23. Clarke, J., Schmidt, M., and Orji, N., Photoresist cross-sectioning with negligible damage using a dual-beam FIB-SEM: A high throughput method for profile imaging. *Journal of Vacuum Science & Technology B* **25**(6): p. 2526-2530, (2007).

24. Dixson, R., et al. CD-AFM reference metrology at NIST and SEMATECH. in *Proc. of SPIE* 5752, (SPIE, 2005). https://doi.org/10.1117/12.601972

25. Dixson, R. and Orji, N., Comparison and uncertainties of standards for critical dimension atomic force microscope tip width calibration. in *Proc. SPIE* 6518, 651816 (SPIE, 2007). https://doi.org/10.1117/12.714032

26. Bunday, B., et al. The coming of age of tilt CD-SEM. in *Proc. SPIE* 6518, 65181S (SPIE, 2007). https://doi.org/10.1117/12.714214

27. Bunday, B., et al. Characterization of CD-SEM metrology for iArF photoresist materials. in *Proc. SPIE* 6922, 69221A (SPIE, 2008). https://doi.org/10.1117/12.774317

28. Martin, Y. and Wickramasinghe, H.K., Method for Imaging Sidewalls by Atomic-Force Microscopy. *Applied Physics Letters*, **64**(19): p. 2498-2500, (1994).

29. Orji, N.G., Dixson, R.G., Bunday, B.D., and Allgair, J.A. Toward accurate feature shape metrology. in *Proc. SPIE* 6922, (SPIE, 2008). https://doi.org/10.1117/12.774426

30. Orji, N.G. and Dixson, R.G., Higher order tip effects in traceable CD-AFM-based linewidth measurements. *Measurement Science and Technology*, **18**(2): p. 448, (2007).

31. Park, B., et al. Application of carbon nanotube probes in a critical dimension atomic force microscope. in *Proc. SPIE* 6518, 651819 (SPIE, 2007). https://doi.org/10.1117/12.712326

Orji, Ndubuisi; Dixson, Ronald; Barnes, Bryan; Silver, Richard.
"Traceability: The Key to Nanomanufacturing."
Paper presented at Sixth International Conference on Precision, Meso, Micro and Nano Engineering, Coimbatore, India. December 11, 2009 - December 12, 2009.

32. Dixson, R.G., Allen, R.A., Guthrie, W.F., and Cresswell, M.W., Traceable calibration of critical-dimension atomic force microscope linewidth measurements with nanometer uncertainty. *Journal of Vacuum Science & Technology B*, **23**(6): p. 3028-3032, (2005).

33. Pennycook, S.J., Scanning Transmission Electron Microscopy: Z-Contrast, in *Electron Microscopy: Principles and Fundamentals* Amelinckx S., et al., Editors.: Weinheim.,(VCH, 1997).

34. Silver, R.M., et al., Scatterfield microscopy for extending the limits of image-based optical metrology. *Applied Optics*, **46**(20): p. 4248-4257, (2007).

35. Gelman, A., *Bayesian data analysis*. Third edition. ed. Chapman & Hall/CRC texts in statistical science. Boca Raton, xiv, 661 pages, (CRC Press, 2003).

36. Silver, R.M., et al. Improving optical measurement accuracy using multi-technique nested uncertainties. in *Proc. SPIE* 7272, 727202 (SPIE, 2009). https://doi.org/10.1117/12.816569

37. Scott, J.H.J., Accuracy issues in chemical and dimensional metrology in the SEM and TEM. *Measurement Science and Technology*, **18**(9): p. 2755-2761, (2007).

38. Orji, N.G., et al., Progress on implementation of a reference measurement system based on a critical-dimension atomic force microscope. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, **6**(2), 023002, (2007).

Examining the True Effectiveness of Loading a Reverberation Chamber

How to Get Your Chamber Consistently Loaded

Jason B. Coder [#1], John M. Ladbury [*2], Christopher L. Holloway [*] and Kate A. Remley [*]

[#] *Department of Electrical Engineering, University of Colorado*
Campus Box 104, P.O. Box 173364, Denver CO 80217 U.S.A.
[1] jason.coder@nist.gov

[*] *National Institute of Standards and Technology*
M.S. 818.02, 325 Broadway, Boulder CO 80305 U.S.A.
[2] john.ladbury@nist.gov

***Abstract*—In this paper we explore how placing the same amount of absorber in different locations within a reverberation chamber can have different loading effects. This difference can have a significant impact on measurement reproducibility, both for measurements in the same chamber and measurements between chambers (i.e., round robin style testing). We begin by discussing some of the theories behind this and show some experimental results from different absorber placements in a reverberation chamber. We conclude with some suggestions to ensure absorber is placed consistently.**

Keywords-Reverberation Chambers, RF Absorber, Power Decay Time, Wireless Deivce Testing

I. INTRODUCTION

As the use of reverberation chambers expands to experiments and measurements involving loaded setups, we explore the realistic effectiveness of absorber loading. Large pieces of absorber are being used routinely in a variety of measurements (for a variety of applications) to lower the Q, decrease the power decay time, and alter the shape of the power delay profile (PDP) [1, 2].

Reverberation chambers are designed to create a consistent test environment with a statistically uniform field throughout the chamber. Many applications have taken advantage of this key principle. For certain applications (i.e., emissions), the original, highly reflective, reverberation chamber is useful. However, applications finding new use of the reverberation chamber are modifying this environment slightly by loading the chamber with RF absorbing materials. An example of this is the use of chambers to test wireless devices [1, 3]. The cell phone industry is beginning to use reverberation chambers for product testing. For the reverberation chambers to be useful for their application, the chamber is loaded to simulate a given insertion loss and power decay time that is representative of a real-world communications channel [4].

Many measurements require the chamber loading to recreate a particular environment. Experiments indicate that the absorber must be within the working volume, but do not specify exact absorber placement. This raises the questions of: Does it matter *where* in the chamber the absorber is placed? Is there an optimal configuration for the antennas and absorber? These are the questions we hope to address.

We address these questions by conducting a series of experiments whereby we measure the relative effectiveness of a fixed quantity of RF absorber placed at different locations within a reverberation chamber. Comparison of measurement results will show the effect of absorber placement within a reverberation chamber.

There are many different parameters we could measure to test our hypothesis. The results could be expressed in terms of power decay time or S-parameters. For this application we choose to display our results in terms of S-parameters. This allows our results to be easily extended to power decay time [1, 2].

II. OUR HYPOTHESIS AND IT'S IMPLICATIONS

We expect small measureable differences as a result of absorber location in the chamber. Absorber will be placed in a corner (between two walls and a floor), along the wall (between one wall and the floor), in the center of the chamber (sitting on floor) and at various positions elevated off the floor.

The volume inside the chamber, where the average field is statistically uniform, is referred to as the "working volume." The boundaries of the working volume are defined to be some distance away from any metallic surfaces in the chamber. The IEC 61000-4-21 standard defines this distance as $\lambda/4$ from any wall, floor, tuner and/or antenna [5]. Other research indicates that field uniformity begins to diminish within $\lambda/2$ of the walls [6]. It is a reasonable assumption that absorber placed within the statistically uniform field of the "working volume" will have a greater loading effect than absorber placed on the floor or next to the walls.

Testing our hypothesis can be accomplished by measuring S_{21}, the insertion loss of the chamber. If our predictions are correct, we should see a change in S_{21} as the absorber is moved around the chamber. The fully exposed absorber in the working volume will absorb more energy than that which is only partially exposed near the floor and walls.

When our measurement results are presented, they will be in the form of $\langle|S_{21}|^2\rangle$. This can easily be thought of in terms of power using the following equation from [7]:

Coder, Jason; Ladbury, John; Holloway, Christopher; Remley, Catherine.
"Examining the True Effectiveness of Loading a Reverberation Chamber: How to Get Your Chamber Consistently Loaded."
Paper presented at 2010 IEEE International Symposium on Electromagnetic Compatibility (EMC), Fort Lauderdale, FL, United States. July 25, 2010 - July 30, 2010.

$$P_{rec} \propto <|S_{21}|^2> \quad (1)$$

where the proportionality constant is P_{trans}. Thinking in terms of power, we predict that the amount of received power will depend on placement of absorber. The power transmitted into the chamber will be constant for each measurement, leaving $<|S_{21}|^2>$ the only variable.

As long as the insertion loss (S_{21}) or similar parameter (power decay time, etc.) is measured before each measurement, any problems with absorber placement can be resolved. We intend to show the outcome of a situation where absorber is moved around *after* the insertion loss is measured. Absorber might be moved around after the insertion loss measurement to make room for a device under test (DUT) or for easier access to parts of the chamber or antennas.

III. Measurement Setup

All measurements were conducted inside a reverberation chamber consisting of two paddles (one floor to ceiling and one wall to wall). The dimensions of the chamber are approximately 4.2 m long, 3.6 m wide and 2.9 m tall. The two paddles are identical in shape and have a width of 0.7 meters. Figure 1 shows the chamber used in our measurements.

Measurements were conducted using a pair of dual-ridged horn antennas (assumed to be identical) mounted on a pair of non-metallic tripods that are 1.3 meters tall. The antennas were connected to a vector network analyzer. Data were acquired at 16,001 discrete points between 800 MHz and 10 GHz and then averaged over 100 discrete paddle steps. When we display our data, we will only show results from 1 GHz to 10 GHz. We discard data below 1 GHz because the antenna mismatch becomes an issue.

The single piece of pyramidal absorber used in these measurements (seen in Figure 1) measures 0.6 m long, 0.6 m wide and 0.6 m tall (to the top of the cones). For the measurements where the absorber is elevated off the floor, styrofoam blocks were used to support the absorber.

Figure 1. Reverberation chamber used in the measurements with the absorber in position 10.

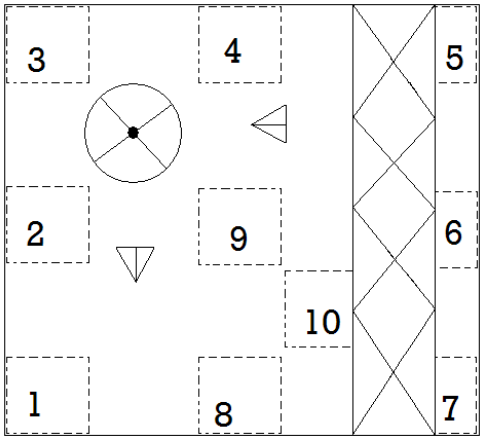


Figure 2. Locations of absorber (dotted lines) in the chamber. One paddle is shown as a circle, and is mounted floor to ceiling while the other paddle outlined (rectangular box) is mounted wall to wall. Positions 5, 6, 7 and 10 are partially under the paddle.

To test our hypothesis that proximity to chamber surfaces affects the loading, absorber was placed in 10 different locations around the reverberation chamber and insertion loss (S_{21}) was measured at each location. The diagram in Figure 2 shows the locations of the absorber in the reverberation chamber. Note that this diagram shows a top down view of the absorber placement.

For positions 1-8, the absorber was placed on the floor of the chamber, up against and touching the wall or corner. In positions 2, 4, 6 and 8, the center of the absorber was placed in the center of the wall. While the absorber was at position 9, data were taken while the absorber was on the floor, and again after the absorber was raised to the height of the antennas (approximately 1.3 meters). This elevated position is referred to as "9A." In position 10, the absorber was at a height of 0.6 meters.

Regardless of the position of absorber, the antennas remained in the same position and orientation (cross-polarized).

IV. Measurement Results and Analysis

At each of the locations shown in Figure 2, S_{21} was measured in phasor form at 100 discrete paddle positions. We computed the squared magnitude of S_{21} at each paddle position, then computed the ensemble average (denoted as < >) to end up with a single $<|S_{21}|^2>$ value at each frequency. In addition to this, we measured the insertion loss of the chamber in an empty configuration as a reference.

We start our look at the measurement results with the reference measurement in Figure 3. Here, the chamber had no absorber in it, that is, there is 0 dB of intentional loading. The losses evident in Figure 3 are due in large part to the wall losses of the chamber. The rippling effect seen at the lower end of the frequency range is due to antenna mismatch. Since the same antennas are used in all of our experiments, any mismatch effects should be constant from measurement to measurement and thus unimportant for our comparisons.

Coder, Jason; Ladbury, John; Holloway, Christopher; Remley, Catherine.
"Examining the True Effectiveness of Loading a Reverberation Chamber: How to Get Your Chamber Consistently Loaded."
Paper presented at 2010 IEEE International Symposium on Electromagnetic Compatibility (EMC), Fort Lauderdale, FL, United States. July 25, 2010 - July 30, 2010.

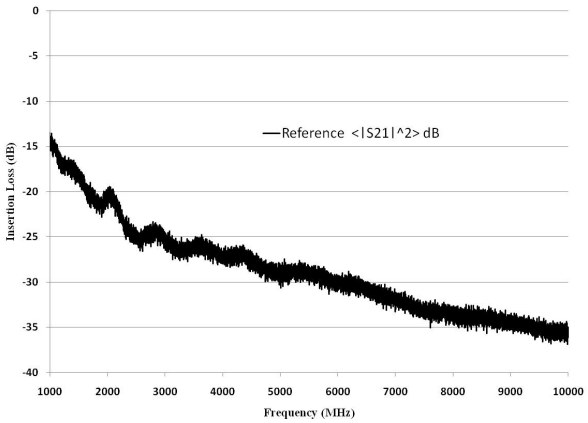


Figure 3. Reference measurement taken with no absorber in the chamber. Insertion loss is shown as $<|S_{21}|^2>$ dB.

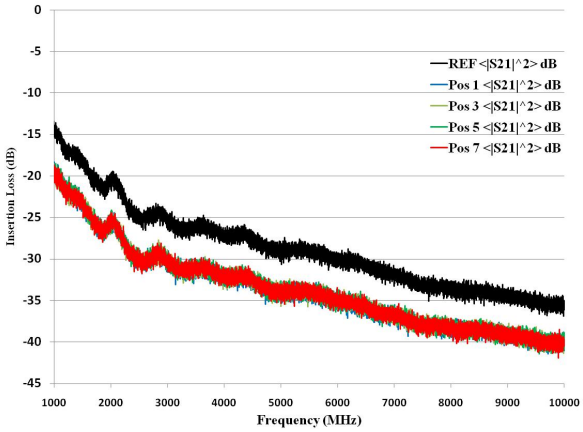


Figure 4. Measured $<|S_{21}|^2>$ dB values for positions 1, 3, 5 and 7 as compared to the reference. No moving average has been applied.

Next, we examine the measurement results for the locations where absorber was placed in the corners of the chamber: positions 1, 3, 5 and 7. In these locations, the absorber should be least effective because it is not fully exposed on 3 sides. Figure 4 shows the four corner locations compared to the reference measurement (Figure 3).

An analysis of Figure 4 shows that each of the four corner absorber positions yields very similar results; so close that the curves are indistinguishable on the plot. To reduce the effects of measurement noise and aid our future analysis, a 101 point moving average is applied to the data. In addition to the moving average, we will compare our results directly to the reference measurement by defining "Absorber effectiveness."Absorber effectiveness is simply the difference between the reference insertion loss (unloaded chamber) and the insertion loss measured at the given location. Mathematically, we define absorber effectiveness as:

$$A.E. = <|S_{21}|^2>_{ref} - <|S_{21}|^2>_{pos.n} \qquad (2)$$

where "*pos.n"* is the given position number.

By calculating A.E. and applying a moving average to the results from positions 1, 3, 5 and 7, we can refine the results of Figure 4 to those of Figure 5 to show that a single piece of absorber in any corner of the chamber adds about 5 dB to the insertion loss measurement.

Figure 5 shows that after reducing the noise (using a moving average) each corner location shows a consistent amount of absorber effectiveness regardless of which corner it is.

The consistency of the corner locations (numbers 1, 3, 5 and 7) can also be seen in the mid-wall locations (numbers 2, 4, 6 and 8). In the mid-wall positions, only two sides of the absorber are not fully exposed. Figure 6 shows these results, with the same moving average applied. Absorber in these locations more effectively loads the chamber than the absorber located in the corners of the chamber.

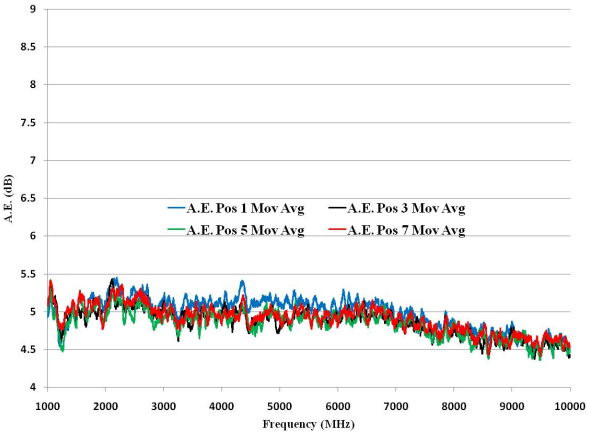


Figure 5. Absorber Effectiveness data for positions 1, 3, 5 and 7. With a 101 point moving average applied.

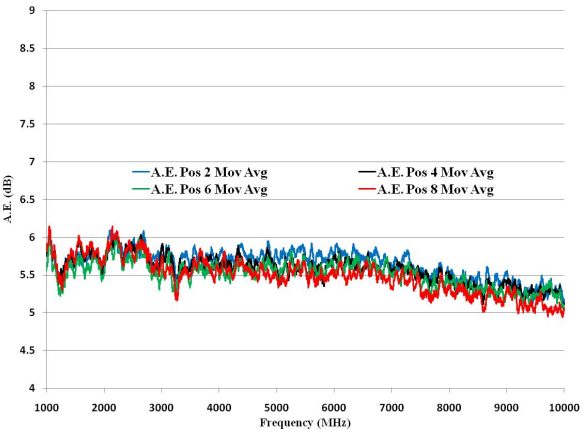


Figure 6. Absorber Effectiveness data for positions 2, 4, 6 and 8 with a 101 point moving average applied.

Coder, Jason; Ladbury, John; Holloway, Christopher; Remley, Catherine.
"Examining the True Effectiveness of Loading a Reverberation Chamber: How to Get Your Chamber Consistently Loaded."
Paper presented at 2010 IEEE International Symposium on Electromagnetic Compatibility (EMC), Fort Lauderdale, FL, United States. July 25, 2010 - July 30, 2010.

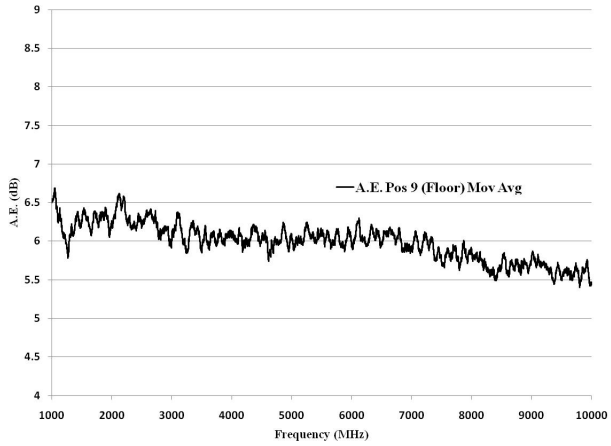


Figure 7. Absorber effectiveness data for position 9 with a 101 point moving average applied.

Next, the absorber was placed in the center of the chamber, on the floor. In this location, the absorber has only one side that is not fully exposed. Therefore, we expect increased absorber effectiveness compared to all previous positions. Figure 7 shows the A.E. of position 9.

In addition to testing a single piece of absorber on the floor at position 9, we did an additional test with the absorber elevated to the height of the antennas (1.3 m). At position 9A, the piece of absorber is totally within the working volume of the chamber and should be fully exposed on all sides. At this position, we would expect that the absorber has reached it's maximum effectiveness. However, as our hypothesis stated, any absorber placed anywhere in the working volume should provide the same amount of effectiveness. To test this, we placed the piece of absorber at position 10 at a height of 0.6 m above the floor. Figure 8 shows that the results from positions 9A and 10 are nearly identical (to within 0.3 dB).

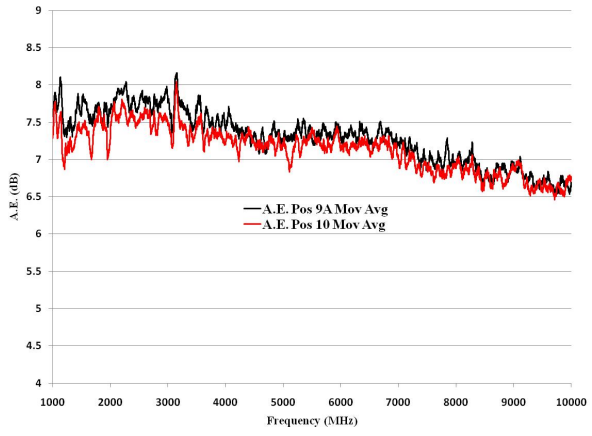


Figure 8. Absorber effectiveness data for positions 9A and 10 with a 101 point moving average applied.

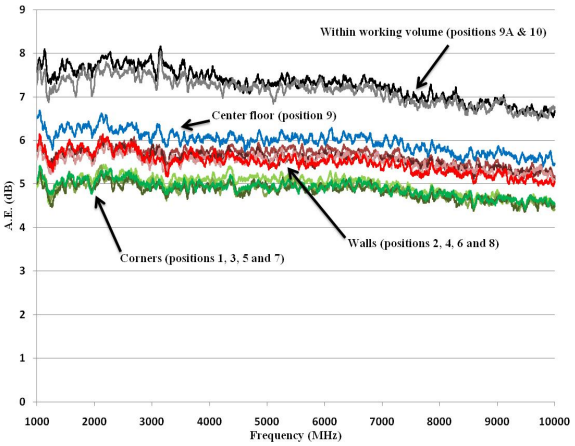


Figure 9. A summary of the absorber effectiveness results showing all 11 configurations tested.

To better compare the differences that arise from placing absorber different locations, Figure 9 shows all 11 configurations on the same plot. On this plot we can see that the absorber placed in the corner of the chamber has a reduced effectiveness of almost 3 dB compared to the most effective location (elevated off the floor). This is most likely due to all sides of the absorber not being fully exposed. Similarly, absorber placed against the wall is approximately 2 dB less effective than if placed well within the working volume (position 9A and position 10).

V. Measurement Uncertainties

Uncertainty for these measurements can be quantified after first identifying the contributing factors. The uncertainty from the statistical processing of the measurements is the largest contributor. This initial uncertainty can be reduced through additional averaging, but cannot be eliminated completely. The second factor to be considered is the system drift as the temperature fluctuates throughout the measurement process.

It is important to note that this uncertainty analysis is restricted to those factors which impact *relative* measurements only. Some uncertainties that would usually affect absolute measurements need not be considered here. For example, uncertainties that manifest themselves as an offset would impact *both* measurements being considered in (2). Uncertainties of this type would not affect the relative measurement because they apply to both measurements equally.

The initial statistical uncertainty can be quantified by the standard deviation of the original measurements, before a moving average is applied (i.e. those shown in Figure 4). With no reference measurement subtracted, the data in each absorber configuration has a standard deviation of approximately 0.4 dB. Computing the absorber effectiveness using (2) increases this uncertainty by approximately 40% ($\sqrt{2}$) to 0.6 dB. To reduce this uncertainty we apply a moving average to the data. This will reduce the uncertainty by the following factor:

Coder, Jason; Ladbury, John; Holloway, Christopher; Remley, Catherine.
"Examining the True Effectiveness of Loading a Reverberation Chamber: How to Get Your Chamber Consistently Loaded."
Paper presented at 2010 IEEE International Symposium on Electromagnetic Compatibility (EMC), Fort Lauderdale, FL, United States. July 25, 2010 - July 30, 2010.

$$U_{stats} = \frac{0.6}{\sqrt{N_f}} \quad (3)$$

where N_f is number of independent frequency samples. This can be determined by calculating the bandwidth of the chamber using the measured Q of the chamber. In the worst case, our measurements have a bandwidth of about 1 MHz. In other words, we must sample at frequency intervals of 1 MHz or more in order to get a statistically independent sample. Given the frequency spacing of our initial measurements, and the size of our moving average window, we have about 50 independent frequency samples in each moving average window. This results in a final U_{stats} of approximately 0.08 dB.

The second contributor to our final uncertainty number is the result of system drift largely due to temperature (U_{temp}). These measurements were taken over three days (two separate system calibrations). Over this period, the system has an estimated drift of about 0.1 dB.

To compute the cumulative uncertainty (also known as the Combined Standard Uncertainty), we combine these two factors by computing the root-sum-of-squares as shown in [8]:

$$U_{comb} = \sqrt{(U_{stats})^2 + (U_{temp})^2} \; . \quad (4)$$

Inserting the terms we determined above, we obtain a combined uncertainty of 0.13 dB. After calculating U_{comb} we can select a coverage factor. We choose a coverage factor of 2, which is common practice [8]. This coverage factor is multiplied with (U_{comb}) to obtain our final measurement uncertainty of 0.26 dB. This value applies to all relative measurements (including absorber effectiveness calculations) presented here.

VI. Discussion

Our measurements to this point seem to coincide with our hypothesis that placing absorber in a corner or against a wall makes it less effective than if it were within the working volume. To more conclusively validate our hypothesis we duplicated the measurements at positions 9A and 10. However, this time we placed a metal plate underneath the piece of absorber. In this configuration, the absorber should be less effective because the metal plate is preventing all sides of the absorber from being exposed. We expect our results to look similar to the difference between positions 9 and 9A; placing a plate under the absorber should simulate the piece of absorber sitting on the floor. However, because the size of the plate is much smaller than the size of the floor, we don't expect the results to be identical. Figure 10 shows the original position 9 and 9A results with the addition of the results from position 9A with a metal plate.

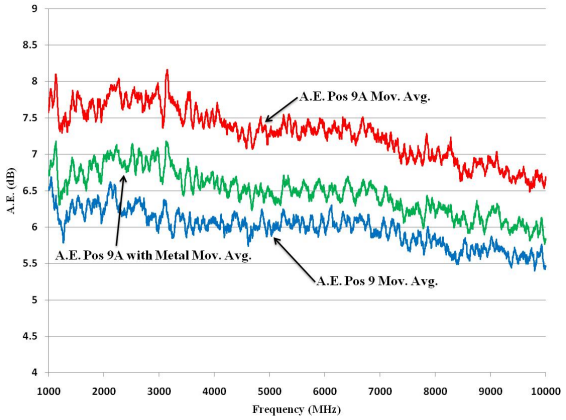


Figure 10. Comparison between Positions 9, 9A and 9A with a metal plate. A 101 point moving average has been applied.

Figure 10 shows that when the metal plate is used, results are within 0.5 dB of position 9. The remaining difference between the two configurations is likely due to the fact that the metal plate used was the exact same size as the bottom of the absorber. When placed on the floor (position 9), the absorber is effectively on a metal plate that is much larger and blocks *all* energy coming from underneath the absorber. With the metal plate, absorber in position 9A can still interact with energy reflected off the floor.

We also placed a metal plate under the absorber in position 10 to show that the result is independent of the absorber location (as long as it is within the working volume). In this configuration, we still expect the result to be close to position 9. Figure 11 shows the original position 9 and 10 data with the addition of the results from placing a metal plate under the absorber in position 10.

As with the metal plate in position 9A, using a metal plate in position 10 yields results that are within 0.5 dB of where we expected to be – equivalent to position 9. The reasoning here is also similar; the floor acts as a much larger metal plate than was actually used in positions 9A and 10.

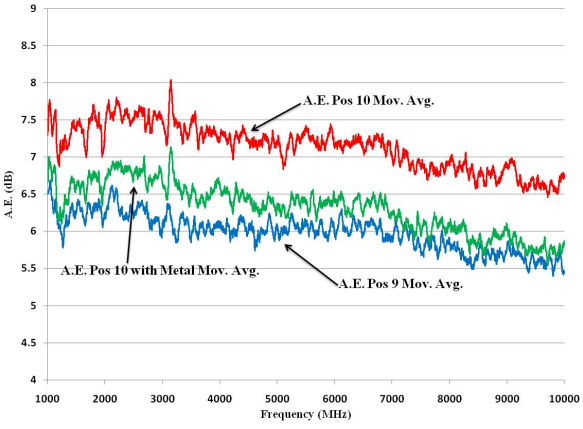


Figure 11. Comparison between Positions 9, 10 and 10 with a metal plate. A 101 point moving average has been applied.

Coder, Jason; Ladbury, John; Holloway, Christopher; Remley, Catherine.
"Examining the True Effectiveness of Loading a Reverberation Chamber: How to Get Your Chamber Consistently Loaded."
Paper presented at 2010 IEEE International Symposium on Electromagnetic Compatibility (EMC), Fort Lauderdale, FL, United States. July 25, 2010 - July 30, 2010.

VII. CONCLUSIONS

We have shown that placing a fixed amount of absorber at different locations inside a reverberation chamber can impact the insertion loss of the chamber (and thus the "effectiveness" of the absorber). By placing a single piece of absorber at 11 different locations we showed that as long as all sides of the absorber are fully exposed and within the working volume of the chamber, maximum effectiveness is achieved. Because of this effect, we encourage others to measure the insertion loss of their chamber – in its test configuration – prior to other measurements and to avoid moving absorber once this measurement has been completed.

REFERENCES

[1] E. Genender, C.L. Holloway, K.A. Remley, J.M. Ladbury, G. Koepke, and H. Garbe, "Simulating the Multipath Channel with a Reverberation Chamber: Application to Bit Error Rate Measurements," Accepted for Publication, IEEE Transactions on Electromagnetic Comp., 2010.

[2] E. Genender, C.L. Holloway, K.A. Remley, J. Ladbury, G. Koepke, and H. Garbe, "Using Reverberation Chambers to Simulate the Power Delay Profiles of a Wireless Environment," EMC Eurpoe 2008 Conference Proceedings, Hamburg Germany.

[3] C. L. Holloway, D. A. Hill, J.M. Ladbury, P. Wilson, G. Koepke and J. Coder, "On the Use of Reverberation Chambers to Simulate a Rician Radio Environment for the Testing of Wireless Devices," IEEE Transactions on Antennas and Propagation, Vol. 54, No. 11, November, 2006.

[4] C. Orlenius, P.S. Kildal, and G. Poilanse, "Measurements of Total Isotropic Sensitivity and Average Fading Sensitivity of CDMA Phones in Reverberation Chamber," IEEE International Symposium on Antennas and Propagation, 2005.

[5] IEC 61000-4-21, "Reverberation Chamber Test Methods," International Electrotechnical Commission, 2001.

[6] D. A. Hill, "Boundary Fields in Reverberation Chambers," IEEE Transactions on Electromagnetic Compatibility, Vol. 47, No. 2, May 2005.

[7] J. Ladbury, G. Koepke and D. Camell, Evaluation of the NASA Langley Research Center mode-stirred chamber facility, National Institue of Standards and Technology, Boulder, CO, 1999, NIST Tech. Note 1508.

[8] B. N. Taylor and C. E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, National Institute of Standards and Technology, Boulder, CO, 1994, NIST Tech. Note 1297.

Coder, Jason; Ladbury, John; Holloway, Christopher; Remley, Catherine.
"Examining the True Effectiveness of Loading a Reverberation Chamber: How to Get Your Chamber Consistently Loaded."
Paper presented at 2010 IEEE International Symposium on Electromagnetic Compatibility (EMC), Fort Lauderdale, FL, United States. July 25, 2010 - July 30, 2010.

Development of Performance Metrics and Test Methods for First Responder Location and Tracking Systems

Francine Amon
Building and Fire Research Lab
NIST
100 Bureau Dr
Gaithersburg MD, 20899
+011 301-975-4913
francine.amon@nist.gov

Camillo Gentile
Information Technology Lab
NIST
100 Bureau Dr
Gaithersburg MD, 20899
+011 301-975-3685
camillo.gentile@nist.gov

Kate A. Remley
Electronics and Electrical Engineering Lab
NIST
325 Broadway St
Boulder CO, 80305
+011 303-497-3672
kate.remley@nist.gov

Categories and Subject Descriptors
D.3.3 [**Programming Languages**]: N/A

General Terms
Measurement, Documentation, Performance, Design, Reliability, Experimentation, Security, Standardization, Verification.

Keywords
First responder, firefighter, evaluation, test methods, location, tracking, test methods

1. EXTENDED ABSTRACT
The need for the development of performance requirements, metrics and test methods for evaluation of location and tracking systems is increasing as these systems begin to become commercially available for use by emergency responders. A working group was recently formed at the 5th Annual PPL Workshop[1] consisting of members of the various emergency responder communities, representatives of government agencies, and location and tracking system manufacturers with plans to collect and share information about their ongoing efforts that have bearing on this effort. The goal of the working group is to develop baseline *performance requirements* for reliable and interoperable indoor location systems for emergency responders and develop *test procedures* to validate vendor systems against those requirements. The final deliverable will be a proposed draft document in acceptable format to be presented to appropriate standards developing organization(s). The following text describes in broad terms the steps by which the draft document will be created.

1.1 Information Collection
Information about past and current activity in the development of performance evaluations of location/tracking systems for first responders will be collected for as many efforts as possible. This serves two main purposes: to learn from the successes and failures of others, and to foster a spirit of cooperation amongst interested parties. This work was begun last year as part of an internally funded NIST project within BFRL. Pertinent information collection items include:

- Thorough literature search on existing solution technologies and manufacturers which offer mature solutions.
- Thorough literature search on existing standards documents and user requirements, including all emergency responder communities.
- Contact key people from government agencies that are involved in or might have an interest in development of performance standards for location and tracking systems (LTS) for first responders.
- Contact emergency responders and manufacturers that are interested in LTS evaluations and invite them to contribute to this effort.

1.2 User Community Input
A very important aspect of this work is to maintain a close relationship with the various LTS user groups so that their requirements and desires are met to the fullest extent possible. The three main user groups are fire fighters, police, and EMS personnel. Other user groups may include the National Guard, the Coast Guard, hazmat, and USAR (technical rescue). In addition to gaining an understanding of the users' needs, information about their operating environments is also essential to design test methods that will ensure the LTS under test will perform adequately in the field. For fire fighters, the LTS must be robust enough to function in a hot, smoke-filled building, regardless of the location technology being used. The input desired from user groups includes:

- Preferred performance requirements: location accuracy, communications reliability, equipment form factor, etc.;
- Description of operation scenarios: building materials, number of floors, number of users, etc.
- Preferred user interface: voice-activated, handheld device, visual display, etc.
- Technical barriers that challenge/prohibit the use of LTS: radio frequency interference, access to building maps, etc.
- Non-technical barriers that challenge/prohibit the use of LTS: lack of resources, political climate, receptiveness of users, etc.

1.3 Roles of Working Group
As this information is collected and questions are answered, the scope of the current project will take shape and reveal new roles for NIST and other stakeholders in the working group. A Steering Group at the PPL Workshop was appointed to provide

 PerMIS'10, September 28-30, 2010, Gaithersburg, MD, USA. ACM 978-1-4503-0290-6-9/28/10

Gentile, Camillo; Amon, Francine; Remley, Catherine.
"Development of Performance Metrics and Test Methods for First Responder Location and Tracking Systems."
Paper presented at PerMIS '10: Performance Metrics for Intelligent Systems Workshop, Baltimore, MD, United States. September 28, 2010 - September 30, 2010.

guidance and order to the efforts of the working group. Some important steps towards achieving the goal of the working group and facilitating the flow of activities are listed below:

- Each stakeholder will provide enough information about their respective activities to allow others to plan their work in a constructive way that may or may not involve collaborations.
- Proceed in the areas that are common among all applications while specific interest groups can work in parallel to develop test methods for the applications with which they are most concerned, with guidance from the Steering Team.
- Leverage as much as possible existing standards for performance requirements and test methods which apply generally to the ruggedness of fire equipment.
- Assign a task group the responsibility for correlating the pieces and assembling them into a cohesive document that can then be channeled to standards developing organizations (SDOs) when the latter are ready to start working on formal standards documents.

1.4 Related Activities at NIST

In the interest of sharing Information about past and current activity in the development of performance evaluations of location/tracking systems for first responders, relevant work at NIST includes:

- The Building and Fire Research Laboratory is compiling a prioritized list of fire scenarios that are important to the fire service with respect to LTS.
- The Building and Fire Research Laboratory is currently developing ground-truth measurement science with which to verify LTS location performance. Future work will incorporate testing of existing LTS in real-scale environments to validate the test methods. The feasibility of using small-scale test facilities for ground-truth measurements and rough-duty ruggedness testing will be explored.
- The Electrical and Electronic Engineering Laboratory is developing an understanding of the way RF signals are attenuated by building materials, reflected and redirected by building surfaces, and how interference from other wireless devices can impact RF signal quality. A measurement campaign will be conducted to characterize RF links in various structure types with respect to time-of-arrival, angle-of-arrival, received-signal-strength, etc.
- The Electrical and Electronic Engineering Laboratory will validate the use of a reduced-scale reverberation chamber in which “virtual” RF signals are generated as a method by which the performance of RF-based, and possibly other types of LTS, is measured.
- The Information Technology Laboratory will develop a computer simulator to extrapolate the tested small-scale performance of actual location/tracking systems to large-scale networks and in fire scenarios in order to validate their performance in realistic deployments.
- The Information Technology Laboratory has been involved in the development and evaluation of many different LTS technologies.
- The Manufacturing Engineering Laboratory is active in the development of standards and guidance documents for search and rescue robots, which involves LTS for certain performance tests included in the suite of requirements for some applications.

2. ACKNOWLEDGMENTS

While this project is just beginning, it would not have gained momentum without financial assistance from the Standards and Technology Directorate of DHS. Also, David Cyganski, Jim Duckworth, and John Orr of Worcester Polytechnic Institute helped considerably by including a session on standards development in their 5th annual Precision Indoor Personnel Location and Tracking for Emergency Responders Workshop. Lastly, our student intern, Adam Kopp’s enthusiasm is infectious and much appreciated.

3. REFERENCES

[1] Fifth Annual Workshop, “Precision Indoor Personnel Location and Tracking for Emergency Responders Workshop”, http://www.wpi.edu/academics/ece/ppl/

Gentile, Camillo; Amon, Francine; Remley, Catherine.
”Development of Performance Metrics and Test Methods for First Responder Location and Tracking Systems.”
Paper presented at PerMIS ’10: Performance Metrics for Intelligent Systems Workshop, Baltimore, MD, United States. September 28, 2010 - September 30, 2010.

A Compact Millimeter-wave Comb Generator for Calibrating Broadband Vector Receivers

Paul D. Hale, Kate A. Remley, Dylan F. Williams, Jeffrey A. Jargon, and C. M. Jack Wang

National Institute of Standards and Technology, Boulder, CO, 80305 USA
e-mail: hale@boulder.nist.gov

***Abstract*—We propose a compact and low-cost waveform source that can be used to calibrate the magnitude and phase response of vector receivers over a bandwidth as large as a few gigahertz. The generator consists of a broadband comb generator followed by a filter and amplifier. As a demonstration, we constructed and calibrated a source suitable for calibrating vector signal analyzers operating between 43 GHz and 46.5 GHz. We then used the source to characterize the response of a commercially available vector signal analyzer. We compared the result with calibrations performed with a general-purpose source.**

***Index Terms*—Calibration, comb generator, oscilloscope, vector signal analyzer, vector signal generator.**

I. Introduction

BROADBAND vector receivers are used to characterize signals that are modulated in phase as well as amplitude such as quadrature amplitude modulation (QAM). A simplified schematic of a vector signal analyzer (VSA, a type of vector receiver) is shown in Fig. 1. The VSA consists of an optional bandpass filter, a downconverter and an analog-to-digital converter. These components cause linear and nonlinear distortion in the signal measurement as demonstrated in, *e.g.*, [1] and [2]. Assuming the instrument is operated in a linear voltage regime, linear distortion dominates the measurement and can be quite significant [1], [2]. If the distortion can be characterized, then digital processing can be used in the VSA to compensate/calibrate the measurement system.

VSA calibration at microwave frequencies is often accomplished by internal calibration routines, however at millimeter-wave frequencies, calibration can be more of an issue. In this work we propose and demonstrate a compact, low-cost waveform generator that can be used to calibrate the response of vector receivers operating in the millimeter-wave frequency regime.

A modulated signal source can be used to calibrate a VSA by assuming that the response of the source is ideal. However, modulated sources can exhibit significant phase errors that increase with modulation bandwidth, as in eg., [3]-[6]. Alternatively the source can be calibrated by use of a sampling oscilloscope and then used to calibrate the vector receiver [5]. However, most currently available general-purpose modulated sources are based on arbitrary waveform generators and frequency converters. These are typically large, expensive, and susceptible to temperature changes and mechanical shock, making them impractical for field applications such as self-calibration of a VSA.

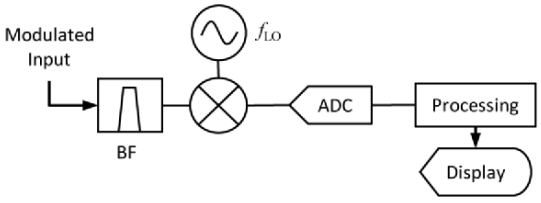


Fig. 1. Simplified schematic of a typical vector signal analyzer consisting of an optional bandpass filter (BF), a downconverter (mixer) and an analog-to-digital converter (ADC). The output of the ADC is digitally processed to form, for example, a constellation diagram or a plot of magnitude and phase verses frequency.

Comb generators are often used for calibrating large-signal network analyzers (LSNAs). Since they can be based on a variety of nonlinear circuits, they are relatively simple, compact, and inexpensive, but are not as versatile as arbitrary waveform generator-based multisine sources. Some comb generators are based on step-recovery diodes and are sensitive to changes in drive power [7], [8] and temperature [9]. Recently, more stable comb generators have been demonstrated, e.g. [8], [10]. These generators produce impulse-like periodic pulses with a broad spectrum.

Because the energy in a train of impulse-like pulses is localized in a very short time interval, the power P_k in the kth tone of the comb is quite low for a given peak voltage V_p. For example, if the pulse repetition frequency is f_0 and the pulse is approximated as rectangular with duration Δ, P_k is given by

$$\begin{aligned} P_k &= \frac{V_p^2 f_0^2}{50\ \Omega}\left(\int_{-\Delta/2}^{\Delta/2}\cos(2\pi k f_0 t)dt\right)^2 \\ &\leq \frac{(V_p f_0 \Delta)^2}{50\ \Omega} \end{aligned}$$

so that $P_k \leq -81$ dBm for typical parameters $V_p = 0.1$ V, $f_0 = 10$ MHz, and $\Delta = 20$ ps. The noise background of a typical VSA is on the order of -100 dBm or more, giving only a 20 dB dynamic range without averaging. (A dynamic range on the order of 40 dB is required to obtain

Hale, Paul; Remley, Catherine; Williams, Dylan; Jargon, Jeffrey; Wang, Chih-Ming.
"A Compact Millimeter-wave Comb Generator for Calibrating Broadband Vector Receivers."
Paper presented at 85th ARFTG Microwave Measurement Conference, Phoenix, AZ, United States. May 22, 2015 - May 22, 2015.

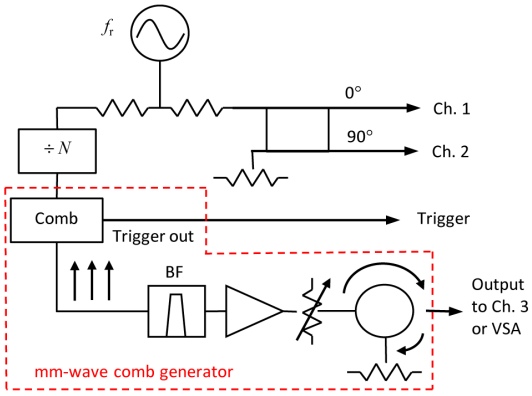


Fig. 2. Schematic drawing of the portable millimeter-wave comb generator and system used to characterize its output with known accuracy. Ports labeled Ch. 1, 2, and trigger refer to connections to a calibrated sampling oscilloscope that is used to calibrate the output of the generator, measured on channel 3. Channels 1 and 2 measure sine waves that are used by the NIST TBC algorithm [13] to correct the oscilloscope timebase. Once calibrated, the comb generator's output can be used to calibrate the linear response of a VSA.

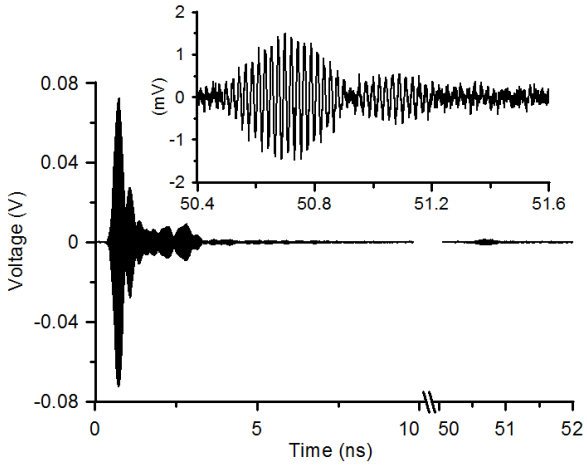


Fig. 3. Measured comb generator waveform.

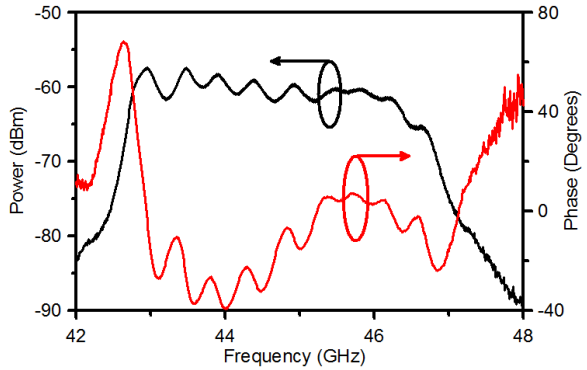


Fig. 4. Frequency domain view of comb generator waveform.

error vector magnitude measurements that are accurate at the $\sim 1\%$ level.) If the pulse duration could be increased, by use of band limiting and/or dispersive elements, while keeping the peak voltage fixed, the power in each tone would be increased, thus improving the signal-to-noise ratio when characterizing a vector receiver. Such an approach was hinted at in [10] and [5]. A source that was designed for a similar application over the 100 MHz to 2 GHz band was described in [11]. In this work we demonstrate a comb-generator-based source, with band-limiting filter, operating between 43 GHz and 46.5 GHz. With this source, other bands up to 67 GHz are realizable.

II. Comb Generator System For millimeter-wave Applications

Our comb-generator system and the apparatus used to characterize its output are shown schematically in Fig. 2. We use commercially available components everywhere in the system. The box labeled "Comb" is a comb generator that is similar to that described in [8] but has been modified to have higher bandwidth and a 1.85 mm coaxial output connector. The output of the comb generator is filtered and temporally broadened by a 43 GHz to 46.5 GHz bandpass filter. Amplification and further filtering are provided by a 20 dB gain amplifier. An optional variable attenuator can be included for studying behavior of the vector receiver under test with respect to the signal power. The isolator is included to make the system source mismatch small and relatively constant when the reflection of the attenuator changes.

Different operating frequency ranges can be chosen by use of different filter/amplifier combinations. The highest useful frequency of our system, 67 GHz, is determined by the bandwidth and coaxial connector of the commercially available comb generator. Higher signal-to-noise ratios can be achieved by higher amplification and/or choosing a narrower bandwidth or a more dispersive filter, such as an all-pass filter [11].

III. Calibration of the Comb Generator System

Calibration of the comb generator system is performed with a calibrated sampling oscilloscope [12] and by using procedures similar to those described in [3], [7], [13]. The system in Fig. 2 and our measurement procedures utilize the publicly available NIST timebase correction algorithm [13], [14]. The signal generator produces sine waves with frequency f_r on the order of 10 GHz used for timebase distortion correction and as the reference for the trigger. The prescaler ($\div N$ block) produces a fast transition that triggers the comb generator, which is measured on channel 3 simultaneously with the sinusoid measurements on channels 1 and 2. Because all of the samplers in the oscilloscope are activated by the same trigger pulse, the timing errors in all the channels in the oscilloscope are nearly identical [14]. The timebase correction algorithm fits the sinusoids measured on channels 1 and 2 and estimates the timing error in their measurement. This estimate is then used to correct the timing error in the measurement of the comb generator.

Portions of a waveform with repetition frequency $f_0 = 10$ MHz are shown in Fig 3. It is important to measure a full period of the waveform because temporally windowing will

Hale, Paul; Remley, Catherine; Williams, Dylan; Jargon, Jeffrey; Wang, Chih-Ming.
"A Compact Millimeter-wave Comb Generator for Calibrating Broadband Vector Receivers."
Paper presented at 85th ARFTG Microwave Measurement Conference, Phoenix, AZ, United States. May 22, 2015 - May 22, 2015.

cause errors if energy is present away from the main pulse, as is the case in our comb generator system (as shown in Fig. 3). In our case, this extra pulse causes ripple in the spectrum with period of 20 MHz and approximately 0.4 dB peak-to-peak magnitude variation (not resolved in Fig. 4) and 0.3 degree peak-to-peak phase variation.

A typical single waveform measurement is timebase corrected by use of the sequential timebase correction algorithm described in [13]. To facilitate further processing steps that use discrete Fourier transforms, the timebase corrected waveform is interpolated onto an evenly spaced temporal grid over an epoch equal to the period of the comb, that is, $1/f_0$. Measurements of the comb are time consuming and subject to temperature drift. The drift error is efficiently corrected by first Fourier transforming the data and then using the phase-alignment algorithm of [15]. After drift correction, the waveforms are transformed back to the time domain and averaged. In our measurements the rms noise of a single waveform is approximately 1 mV and is reduced to 0.1 mV by averaging 100 measurements. Finally, the averaged measurements are corrected for the response of the oscilloscope and impedance mismatch, as described in [16], to obtain the response of the comb generator. Note that, while the samplers are calibrated by electro-optic sampling techniques through a National Metrology Institute, all other steps described above may be implemented by users.

IV. Calibration of a VSA and Comparison With Results Obtained with Other Sources

We used our comb generator system to calibrate the response of a commercially available VSA over a bandwidth of 120 MHz centered at 45 GHz. For comparison, we also synthesized two different multisines with 10 MHz tone spacing centered at 45 GHz [4]. These multisines were calibrated with the oscilloscope in the same way as with the comb generator system and had peak voltages that were comparable to that of the comb generator system. The first multisine had a (nominally) constant phase with respect to frequency and the second had Schroeder phase [17] characteristic.

The VSA was configured to acquire an integer number of cycles of the comb [18]. The input range was set about 4 dB below saturation. Multiple measurements of all signals were acquired, aligned using the technique of [15] and then averaged. Measurements of the comb generator system were aligned to a constant phase of zero.

Results of the measurements are shown in Fig. 5. The phase of the VSA measurements were compared to those measured by the calibrated oscilloscope. Agreement in the characterization of the VSA's phase error indicates that the portable comb system is performing on par with the rack of equipment that constitutes the AWG-based source. To quantify this, we used a paired t-test [19] to compare the mean measured phase $\overline{\phi}_{\mathrm{MS}}(f_j) = (\phi_{\mathrm{C}}(f_j) + \phi_{\mathrm{S}}(f_j))/2$ of the two multisine signals with the comb phase $\phi_{\mathrm{comb}}(f_j)$. The difference $d_j = \overline{\phi}_{\mathrm{MS}}(f_j) - \phi_{\mathrm{comb}}(f_j)$ averaged over the 13 frequencies is $\overline{d} = 3 \times 10^{-5}$ and its standard deviation is $s_{\overline{d}} = 0.042$. The test statistic $\overline{d}/s_{\overline{d}} = 7.5 \times 10^{-4}$ produces

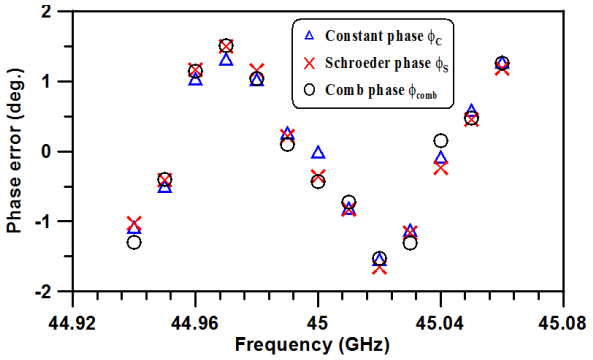


Fig. 5. Difference between phase measured with oscilloscope and phase measured with VSA based on measurements with the comb (○) and multisines with fixed phase (△) and Schroeder phase (×).

a p-value of 0.999. Therefore, based on the data, we find no significant difference between the multisine- and comb generator-based calibrations.

V. Conclusion

We presented a straightforward measurement technique for calibrating vector signal analyzers at millimeter-wave frequencies by use of a calibrated comb generator. The method uses off-the-shelf parts and is portable, allowing users to conduct these calibrations in their own labs, without use of cumbersome, expensive AWG-based sources. The results presented here demonstrate that our prototype millimeter-wave comb generator can be used for calibrating VSAs without a significant degradation in the calibration of the phase response.

VI. Acknowledgements

This work was partially supported by the Defense Advanced Research Projects Agency's Efficient Linearized All-Silicon Transmitter ICs (ELASTx) Program. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Department of Defense.

References

[1] C. J. Clark, A. A. Moulthrop, M. S. Muha, and C. P. Silva, "Transmission response measurements of frequency-translating devices using a vector network analyzer", *IEEE Trans. Microw. Theory Tech.*, vol. 44, no. 12, pp.2724 -2737, Dec. 1996.

[2] D. F. Williams, H Khenissi, F. Ndagijimana, K. A. Remley, J. P. Dunsmore, P. D. Hale, C. M. Wang, and T. S. Clement, "Sampling-oscilloscope measurement of a microwave mixer with single-digit phase accuracy," *IEEE Trans. Microwave Theory Tech.*, Vol. 54, pp. 1210-1217, Mar. 2006.

[3] K. A. Remley, P.D. Hale, D.I. Bergman, D. Keenan,"Comparison of multisine measurements from instrumentation capable of nonlinear system characterization," *66th ARFTG Conf. Dig.*, Dec. 2005, pp. 34-43.

[4] K. A. Remley, P. D. Hale,D. F. Williams, J. Wang, "A precision millimeter-wave modulated-signal source", *2013 IEEE MTT-S Int. Microwave Symp.*, 2013.

[5] D. A. Humphreys, M. R. Harper, and M. Salter, "Traceable calibration of vector signal analyzers", *75th ARFTG Microwave Meas. Conf.*, 2010.

Hale, Paul; Remley, Catherine; Williams, Dylan; Jargon, Jeffrey; Wang, Chih-Ming.
"A Compact Millimeter-wave Comb Generator for Calibrating Broadband Vector Receivers."
Paper presented at 85th ARFTG Microwave Measurement Conference, Phoenix, AZ, United States. May 22, 2015 - May 22, 2015.

[6] D. Rabijns, W. Van Moer, and G. Vandersteen, "Using multisines to measure state-of-the-art analog-to-digital converters," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 3, pp. 1012-1017, June 2006.
[7] H. C. Reader, D. F. Williams, P. D. Hale, and T. S. Clement, "Characterization of a 50 GHz comb generator," *IEEE Trans. Microwave Theory Tech.*, vol. 56, no. 2, pp. 515-521, Feb. 2008.
[8] D. Gunyan and Y.-P. Teoh, "Characterization of active harmonic phase standard with improved characteristics for nonlinear vector network analyzer calibration," *2008 IEEE MTT-S Int. Microwave Symp.*, pp. 73-79, 2008.
[9] J. A Jargon, J. D. Splett, D. F. Vecchia, and D. C. DeGroot, "An empirical model for the warm-up drift of a commercial harmonic phase standard," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 3, pp. 931-937, June 2007.
[10] D. B. Gunyan and J. B. Scott, U.S. Patent No. 7,423,470 B2, Sept. 9, 2008.
[11] T. Van der Broeck, R. Pintelon, and A. Barel, "Design of a microwave multisine source using allpass functions estimated in the Richards domain," *IEEE Trans. Instrum. Meas.*, vol. 43, no. 5, pp. 753-757, Oct. 1994.
[12] T. S. Clement, P. D. Hale, D. F. Williams, C. M. Wang, A. Dienstfrey, and D. A. Keenan, "Calibration of sampling oscilloscopes with high-speed photodiodes," *IEEE Trans. Microwave Theory Tech.*, vol. 54, pp. 3173-3181 (Aug. 2006).
[13] C. M. Wang, P. D. Hale, J. A. Jargon, D. F. Williams, and K. A. Remley, "Sequential estimation of timebase corrections for an arbitrarily long waveform," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 10, pp. 2689-2694, Oct. 2012.
[14] P. D. Hale, C. M. Wang, D. F. Williams, K. A. Remley, and J. Wepman, "Compensation of random and systematic timing errors in sampling oscilloscopes," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 6, pp. 2146-2154, Dec. 2006.
[15] K. A. Remley, D. F. Williams, D. Schreurs, G. Loglio, and A. Cidronali, "Phase detrending for measured multisine signals, *61st ARFTG Conf. Dig.*, June 2003, pp. 73-83.
[16] D. F. Williams, T. S. Clement, P. D. Hale, and A. Dienstfrey,"Terminology for High-Speed Sampling-Oscilloscope Calibration," *68th ARFTG Conference Digest, Boulder, CO, Nov. 30-Dec. 1, 2006*.
[17] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, New York, NY: IEEE Press, 2001.
[18] M. D. McKinley, K. A. Remley, M. Myslinski, and J. S. Kenney,"Eliminating FFT artifacts in vector signal analyzer spectra," *Microwave Journal*, vol. 49, no. 10, pp. 156-164, Oct. 2006.
[19] Peter W. M. John, *Statistical Methods in Engineering and Quality Assurance*, John Wiley & Sons, 1990.

Hale, Paul; Remley, Catherine; Williams, Dylan; Jargon, Jeffrey; Wang, Chih-Ming.
"A Compact Millimeter-wave Comb Generator for Calibrating Broadband Vector Receivers."
Paper presented at 85th ARFTG Microwave Measurement Conference, Phoenix, AZ, United States. May 22, 2015 - May 22, 2015.

XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 - September 4, 2015, Prague, Czech Republic

PROGRESS ON MAGNETIC SUSPENSION FOR THE NIST VACUUM-TO-AIR MASS DISSEMINATION SYSTEM

Corey Stambaugh [1,*], *Edward Mulhern* [1], *Eric Benck* [1], *Zeina Kubarych* [1], *and Patrick Abbott* [1]

[1] National Institute of Standards and Technology, Gaithersburg, U.S.A., *corey.stambaugh@nist.gov

Abstract - The redefined kilogram will be realized in vacuum and NIST is developing a magnetic suspension system to disseminate the standard to air. This paper details the progress to characterize and improve the performance of the magnetic suspension portion of the system. Effort has been made to determine the magnetic field within the system through simulation and experiment. To increase stability interferometry is being integrated into the system.

Keywords: mass, balance, magnet, suspension, interferometry

1. INTRODUCTION

In 2018, the SI unit of mass, the kilogram, is expected to be redefined so that its magnitude is set by fixing the value of the Planck constant. The reasons behind this change [1] include the measured difference in mass between the international prototype kilogram and the other official copies produced during the same time period [2]. Despite this redefinition, the general process of mass dissemination will remain largely the same and thus continued use of mass comparators and artifacts is necessary. However, the primary realization of the new kilogram will take place in vacuum, therefore a method of vacuum-to-air transfer is needed.

The common method of vacuum-to-air mass comparison is to use sorption standards to develop empirical models to account for mass changes that air introduces when the test mass is removed from vacuum [3]. At the National Institute of Standards and Technology a different approach is being pursued. Here, the standard artifact, kept in vacuum, is directly compared, using the same high precision balance, to a test mass in air. The mass comparison is carried out by coupling the two systems through a magnetic suspension method [4]. This approach eliminates the need to correct the test mass value to account for environmental sorption effects.

Currently, the magnetic suspension system (MSS) is going through an upgrade as the entire setup is moved into a new laboratory. In this paper we discuss the work being done during this time to characterize and improve the overall performance of the MSS. The MSSs discussed here encompass two systems. The first, which will be referred to as the real MSS, represents the final upgraded system. The second, referred to as the test MSS, is a prototype built for testing improvements to the magnet suspension before integration into the real MSS. In general the magnetic suspension systems can be broken down into several key components: (1) the actuator which consists of the electromagnetic coil and permanent magnets, (2) the sensor which consists of a Hall probe and interferometer, and (3) the feedback electronics and software. Through these studies we aim to improve the system stability and increase the overall precision of the mass comparison measurements.

2. MAGNETIC SUSPENSION SYSTEM

Figure 1 shows a schematic of the magnetic suspension system, while Fig. 2a identifies the components of the magnet assembly used in the real MSS. The entire assembly is housed in the upper and lower chambers of an aluminum vacuum vessel. A precision mass comparator is located in the evacuated upper chamber; its mass pan and the upper magnet assembly are hung from the comparator. The upper magnet assembly consists of a rare-earth cylinder magnet with a soft-iron cylinder placed above it; the iron serves to enhance the field. Wrapped around the permanent magnet is a multi-turn coil. Surrounding the magnet are two layers of mu-metal shielding. This reduces the spatial extent of the magnet's field and limits its interaction with any surrounding magnetic materials. The lower chamber at atmospheric pressure contains a second magnet assembly with a mass plate connected below it. The lower magnet assembly is similar to the upper except there is no coil. Embedded in the flange separating the upper and lower chambers is a Hall sensor. In the upgraded system an optical measurement sensor will be attached directly below the lower magnet assembly.

2.1. Mass Measurement

Using this system, mass comparison could be carried out in the following way. First, a reference mass, calibrated using the Watt balance [5], is placed on the mass pan positioned above the magnet assembly in the upper chamber. The lower magnet assembly, without the test mass, is then suspended and a reading of the mass using the balance is carried out. Next the magnetic suspension is stopped, the reference mass is removed from the upper pan and a test mass is placed on the lower plate, then the lower assembly is re-suspended and a measurement of the test mass is made. The test mass is then determined to be the calibrated mass of the standard plus the difference in the two mass measurements. Additional adjustments are made to account for the buoyancy correction due to air and the gravitational change due to the differing heights of the measured masses.

Stambaugh, Corey; Mulhern, Edward; Benck, Eric; Kubarych, Zeina; Abbott, Patrick.
"progress On Magnetic Suspension for the NIST Vacuum-To-Air Mass Dissemination System."
Paper presented at 92nd ARFTG Microwave Measurement Conference, Prague, Czechia. August 30, 2015 - September 4, 2015.

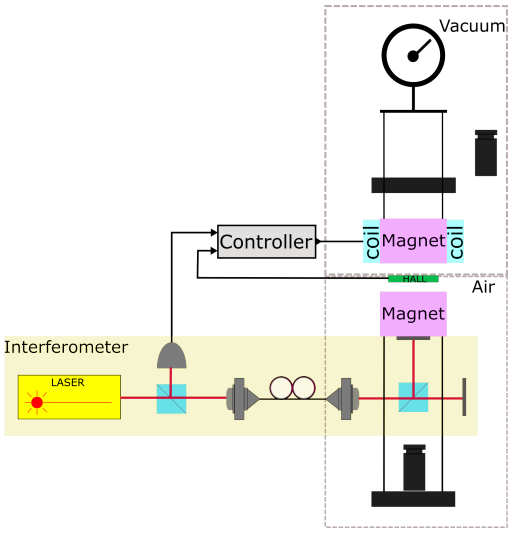


Fig. 1. Schematic of the magnetic suspension system for vacuum-to-air mass comparison. The reference mass is housed in the upper vacuum chamber along with the mass balance. The test mass is placed in the lower chamber. The relative separation of the upper and lower assembly is monitored by a Hall probe. In the new design an interferometric signal will be used for monitoring sub-micron motion. These two signals are fed into the servo controller; the output drives the electromagnetic coil stabilizing the suspended assembly and improving the precision of the mass measurement.

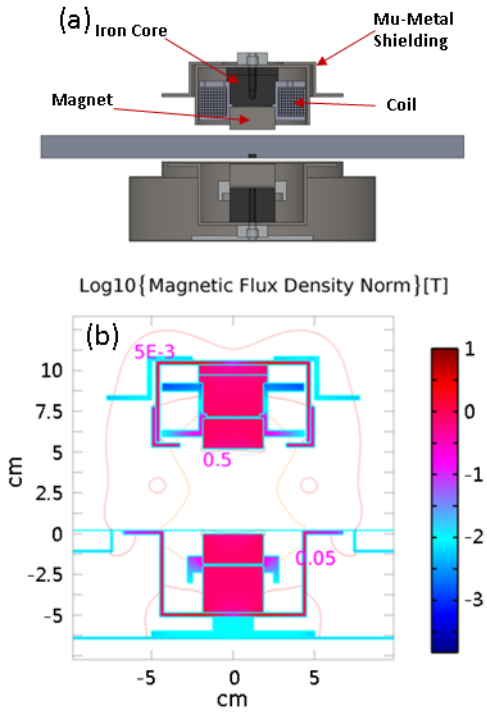


Fig. 2. (a) Cross-sectional diagram of the upper and lower assemblies. (b) Magnetic field strength distribution around the upper and lower assemblies. The model shows the effectiveness of the mu-metal shielding in reducing the magnetic field outside of the assemblies.

3. ACTUATING

3.1. Electromagnetic Coil

The electromagnet coil used in the experiment consists of a multi-turn coil. The coil is driven by an H-bridge circuit that is controlled by a pulse-width modulated signal. In the ideal case the lower assembly will be suspended such that the average current flowing through the coil is zero. However this is not always the case, and since the coil is placed in vacuum heat dissipation becomes a concern. In the original MSS, which used a NdFeB (neodymium) magnet, local heating by the coil may have lowered the magnetization of the permanent magnet, weakening the force between the upper and lower assembly. In the upgraded real MSS we will switch to SmCo (samarium−cobalt) magnets which have a Curie temperature of approximately 2.5 times that of NdFeB magnets.

3.2. Permanent Magnet

Understanding the magnetic field distribution and strength around the magnet assemblies is essential to ensure the field does not couple to some external magnetic material. Additionally, the field gradient along the axial direction determines the magnetic force between the magnets and thus the maximum load. The 2009 paper by Jabbour et al. [6] examined the magnetic field distribution of the original system. However, the finite element modeling (FEM) was not compared to real-world data and several changes have been made to the system, including the switch to SmCo magnets. A 2D cross-section of the current design is shown in Fig. 2a. The modeling of the field, Fig. 2b, is carried out using the commercial software COMSOL[1]. To facilitate an understanding of the modeling and to validate its usage we started with a simple system that could be checked against analytic solutions and simulations and then moved to the full system.

Cylinder Magnet

The core component of the magnetic suspension system is the cylindrical permanent magnet used to produce the dc magnetic field. The majority of the pulling force used in the suspension results from this field. In the current

[1]Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Stambaugh, Corey; Mulhern, Edward; Benck, Eric; Kubarych, Zeina; Abbott, Patrick.
"progress On Magnetic Suspension for the NIST Vacuum-To-Air Mass Dissemination System."
Paper presented at 92nd ARFTG Microwave Measurement Conference, Prague, Czechia. August 30, 2015 - September 4, 2015.

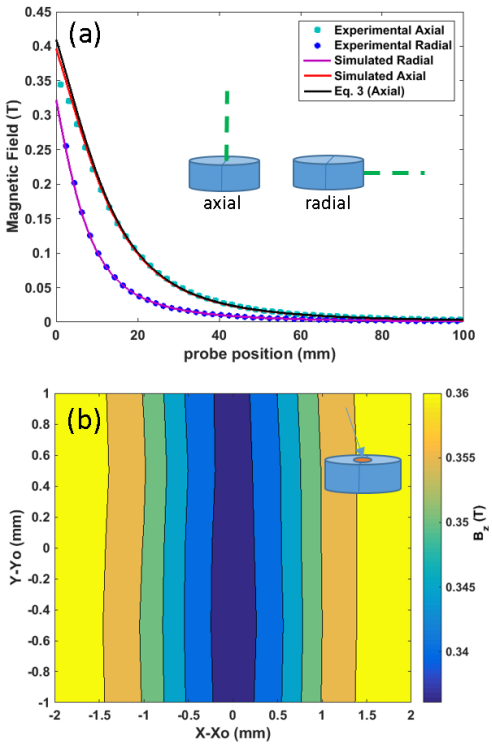


Fig. 3. (a) Magnetic field strength measured along the axial and radial direction. (b) Magnetic field strength measured radially about the center of the magnet, 1.05 mm off the surface. Here a lack of axial symmetry can be clearly identified.

design, the magnet has a height of 19 mm and a diameter of 38 mm. We focus here on a SmCo magnet with a residual flux density of $B_{res} = 1.16$ T.

A common approach to deriving an analytic expression, in the case of no current density, is to use a scalar magnetic potential:

$$\Psi = \int_V \frac{\rho_m(\vec{r})dr}{4\pi\mu_0|\vec{r}-\vec{r}'|}, \tag{1}$$

where $\rho_m(\vec{r})$ is the magnetic charge density and μ_0 is permeability of free space. The magnetic field can be determined by inserting Ψ into

$$\vec{H} = \vec{\nabla}\Psi \tag{2}$$

and using $B_z = \mu_0 H_z$. The resulting magnetic field strength is

$$B_z = \frac{B_{res}}{2}\left(\frac{z-d/2}{\sqrt{R^2+(z-d/2)^2}} - \frac{z+d/2}{\sqrt{R^2+(z+d/2)^2}}\right). \tag{3}$$

Here R and d are, respectively, the radius and height of the cylinder and z is the distance the field is measured from the center of the magnet.

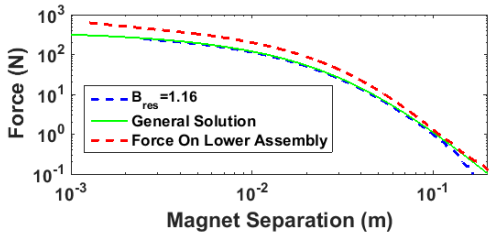


Fig. 4. Comparison of FEM calculation (blue dashed) for two cylindrical SmCo magnets to analytic expression (green line), (5), for the magnetic force. The red dashed line is the FEM calculation for the force on the entire lower assembly.

Using Ampere's law we carry out an FEM of the system to determine the field. To decrease computation time, we take advantage of the 2D axial symmetry of the model. Within the geometry we also add a small fillet to the corners of the magnets, while this does not impact the resultant field, it can have a substantial impact on the force calculation. The material property is chosen to be iron, however the relative magnetic permeability and residual flux density are set to those of SmCo, namely $\rho = 1.05$ and $B_{res} = 1.16$ T, respectively. This choice captures the essence of the important magnetic properties of the material.

A comparison of the results is shown in Fig. 3a, where strong agreement is seen except near the surface where the measured field decreases. The radial measurements in Fig. 3a, also show strong agreement. Finally, in Fig. 3b the field around the central axis at a fixed axial separation is measured. While for true axial symmetry a radial distribution would be expected, here, a clear breaking of this symmetry is found. This deviation is the same decrease seen in Fig. 3a, and results from a small residual magnetization pointing inward along the x-axis, toward the center that resulted from the manufacturing process. This unexpected result should not negatively impact performance, and may help dampen rotation of the lower magnet about its axial axis while suspended.

Magnetic Force

A closed-form solution for the force between two cylindrical magnets can be derived [7]. The general expression is

$$F_z = \frac{\partial E}{\partial z} = 2\pi\mu_0 M^2 R^3 \frac{\partial J_d}{\partial z}, \tag{4}$$

from which a general solution can be found:

$$F_z = -2\pi K_d R^2 \sum_{i,j=-1}^{1} i\cdot j\cdot A_{11}^0(\zeta + i\tau_1 + j\tau_2, 1, 1), \tag{5}$$

where $K_d = \mu_0 M^2/2$ and

$$A_{11}^0(\omega,1,1) = \frac{\omega}{\pi k_1}E(k_1^2) - \frac{(2+0.5\omega^2)k_1\omega}{2\pi}K(k_1^2) + \frac{1}{2}. \tag{6}$$

Stambaugh, Corey; Mulhern, Edward; Benck, Eric; Kubarych, Zeina; Abbott, Patrick.
"progress On Magnetic Suspension for the NIST Vacuum-To-Air Mass Dissemination System."
Paper presented at 92nd ARFTG Microwave Measurement Conference, Prague, Czechia. August 30, 2015 - September 4, 2015.

K and E are complete elliptic integrals of the first and second kind and $k_1^2 = 4/(4+\omega^2)$. As seen in Fig. 4 the general solution shows excellent agreement for separations up to 10 cm; beyond this distance the FEM breaks down. An expression for the force acting on the entire lower assembly is more difficult to derive, so the FEM is necessary to determine the force.

4. SENSING

4.1. Hall probe

A Hall probe is placed between the two magnet assemblies to monitor the magnetic field. The field indicates the relative separation of the two magnets and thus provides a good signal for the servo to stabilize. As part of the effort to improve the overall system we will move to a Hall probe with a lower noise floor. As is often the case during such a testing period we have also identified limitations in the current A/D conversion of the signal. We are working to implement fixes that will lead to a decrease in the minimum resolution of the measured signal.

4.2. Interferometer

Given the measurements made of the Hall probe signal and the magnetic field gradient computed based on simulation, a lower limit on the position resolution and ultimate stability of the suspended mass can be determined. To further decrease this value, efforts have been made to design and implement an interferometer to detect fine motion of the suspended assembly.

The simplified approach, as shown in Fig. 1, involves configuring the interferometer such that its measurement arm is directly below the suspended assembly. In this way, vertical motion can be detected. The interferometry is based on a heterodyne measurement. Two orthogonally polarized beams of light, shifted in frequency by 2.8 MHz, are injected into a fiber and passed into the lower chamber. The beams are then sent through a polarizing beam splitter. One polarization becomes the reference arm of the interferometer while the other is the measurement arm that is bounced off the suspended assembly. The beams are recombined and sent back through the fiber where they are passed through a polarizer set at 45° and then measured using a photo-detector. The electrical response is then demodulated at the unperturbed frequency difference. The movement of the hanging assembly Doppler shifts the measurement beam, which leads to a frequency difference. This difference can be converted to determine both the velocity and the displacement of the suspended lower magnetic assembly.

5. CONTROL

To control and stabilize the suspended mass, feedback is used. In the previous generation the MSS feedback was carried out by a PID loop that was implemented and run directly from a desktop computer. Typical computer latency led to jitter in the feedback loop rate, this in turn led to instabilities in the suspension of the mass. We have since moved the entire feedback system to a field programmable gate array (FPGA). The FPGA approach provides a deterministic loop rate from which to implement the feedback. We currently run at a rate of 100 kHz, an order of magnitude faster than the average loop rate we could previously obtain. Additionally, because of the ease of altering code on the FPGA, as opposed to a pure analog approach, the use of multiple control signals (Hall sensor, interferometric velocity and displacement) is simplified. Of course, we have to confront resolution issues resulting from the limited number of bits available for digitization.

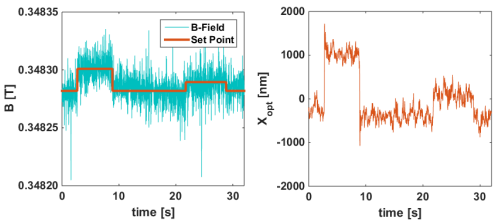


Fig. 5. Control of suspended mass by varying set point. Using interferometer coupled with the Hall sensor vertical displacement less than 500 nm can be actuated and detected.

Using the test MSS we have implemented the new FPGA code for system control. This in conjunction with a 24-bit ADC and the newly added interferometric signal has provided a high level of control within the system. In Fig. 5, we show changes to the measured magnetic field as well as the optically measured displacement of the suspended mass pan, resulting from small changes to the set point. Within the level of stability currently achieved with the test MSS we can clearly see sub-micron motion. At this time the mass stability is below the resolution (1 mg) of the comparator used in the test MSS. As such, no measurable change in mass is seen during the controlled displacements shown in Fig. 5.

6. CONCLUSIONS

NIST is actively working to develop a mise-en-pratique, a set of instructions for the definition of a unit that ensures it can be realized at the highest level of precision, for the redefinition of the kilogram. A critical component of this is the mass comparison of a mass standard in vacuum to a test mass in air. To that end, recent efforts have been focused on the magnetic suspension system. The system is currently being re-assembled in a newly renovated laboratory. Building on our experiences from the previous system and incorporating results from the tests and improvements described within this paper should provide a system that meets the demands of the mass community, i.e. mass comparison measurements with a standard uncertainty for 1 kg of less than 20×10^{-9} kg.

Stambaugh, Corey; Mulhern, Edward; Benck, Eric; Kubarych, Zeina; Abbott, Patrick.
"progress On Magnetic Suspension for the NIST Vacuum-To-Air Mass Dissemination System."
Paper presented at 92nd ARFTG Microwave Measurement Conference, Prague, Czechia. August 30, 2015 - September 4, 2015.

REFERENCES

[1] I. M. Mills, P. J. Mohr, T. J. Quinn, B. N. Taylor, and E. R. Williams, "Redefinition of the kilogram: a decision whose time has come," *Metrologia*, vol. 42, no. 2, p. 71, 2005.

[2] G. Girard, "The third periodic verification of national prototypes of the kilogram (1988-1992)," *Metrologia*, vol. 31, no. 4, p. 317, 1994.

[3] A. Picard and H. Fang, "Methods to determine water vapour sorption on mass standards," *Metrologia*, vol. 41, no. 4, p. 333, 2004.

[4] F. T. Holmes, "Axial magnetic suspensions," *Review of Scientific Instruments*, vol. 8, no. 11, pp. 444–447, 1937.

[5] S. Schlamminger, D. Haddad, F. Seifert, L. S. Chao, D. B. Newell, R. Liu, R. L. Steiner, and J. R. Pratt, "Determination of the planck constant using a watt balance with a superconducting magnet system at the national institute of standards and technology," *Metrologia*, vol. 51, no. 2, p. S15, 2014.

[6] Z. J. Jabbour, P. Abbott, E. Williams, R. Liu, and V. Lee, "Linking air and vacuum mass measurement by magnetic levitation," *Metrologia*, vol. 46, no. 3, p. 339, 2009.

[7] D. Vokoun, M. Beleggia, L. Heller, and P. ittner, "Magnetostatic interactions and forces between cylindrical permanent magnets," *Journal of Magnetism and Magnetic Materials*, vol. 321, no. 22, pp. 3758 – 3763, 2009.

Stambaugh, Corey; Mulhern, Edward; Benck, Eric; Kubarych, Zeina; Abbott, Patrick.
"progress On Magnetic Suspension for the NIST Vacuum-To-Air Mass Dissemination System."
Paper presented at 92nd ARFTG Microwave Measurement Conference, Prague, Czechia. August 30, 2015 - September 4, 2015.

Analytical Transmission Scanning Electron Microscopy: Extending the Capabilities of a Conventional SEM Using an Off-the-Shelf Transmission Detector*

Jason Holm[1] and Robert Keller[1].

[1.] National Institute of Standards and Technology, Materials Measurement Lab., Boulder, CO 80305

Transmission-mode scanning electron microscopy techniques are seeing increased use in both materials science and biological applications [1]. One reason is that recently available off-the-shelf transmission detectors that can be implemented into nearly any scanning electron microscope (SEM) are enabling imaging modes not normally associated with conventional SEMs. For example, some modular units include bright and dark field transmission detectors, and one modular unit includes a high angle annular dark field (HAADF) detector with which elemental contrast can be observed.

Although modular transmission detectors have demonstrated much utility [1-3], their full capacity has not yet been realized. One drawback is the inability to easily control the detector collection angle, and therefore select which electrons will be used to form images. Unlike scanning transmission electron microscopes (STEMs), conventional SEMs do not have post-specimen projector lenses that can be used to modify the collection angle. This means either (1) the specimen-to-detector distance must be adjusted to change the collection angle, or (2) an appropriate mask/aperture must be used to restrict the scattered electrons to a particular region of the detector corresponding to the desired collection angle. Furthermore, SEM transmission detectors with built-in specimen holders do not allow the sample position or orientation to be changed with respect to the detector or incident electron beam, and commercially available detectors with separate holders allow very little specimen positional control. (For example, see Fig. 1 in [2].) Inserting masks/apertures between the specimen and detector in such cases is challenging since very little space is available, and aligning an aperture with the detector and the optic axis is cumbersome at best.

This contribution demonstrates a method to extend the analytical capacity of existing off-the-shelf modular detectors by implementation of (1) a new sample holder, and (2) a simple, inexpensive aperture system that can be used to select electrons scattered to any angle intercepted by the detector. It will be shown for one particular transmission detector that bright field acceptance angles are not limited to those dictated by the intended bright field detector, and that acceptance angles are easily controllable and modifiable over a wide range, from 0 mrad to > 1350 mrad. Implementation of other transmission imaging modes including annular bright field (ABF) [4], low angle annular dark field (LAADF) [5], and thin annular detector (TAD) schemes [6] is also possible. In particular, this work shows how the two developments can enable HAADF transmission imaging in a 15 year old SEM.

To demonstrate the utility of a cantilevered, clamp-based sample holder and aperture system, Figure 1 shows an interior view of the SEM chamber with a transmission detector and EDS detector inserted. An annular aperture (not directly visible in the chamber image) was placed between the transmission detector and the sample. With this aperture in place, and the sample positioned approximately 6 mm above the detector, the dark field inner acceptance angle was approximately 250 mrad. Figure 2 shows dark field images of several 30 nm diameter Au and TiO_2 nanoparticles on a carbon substrate recorded under these conditions. The image intensity differences for different materials are easily discernible in Fig 2a. Figure 2b shows an X-ray map of the sample, demonstrating that the most intense regions can

lm, Jason; Keller, Robert.
"Analytical Transmission Scanning Electron Microscopy: Extending the Capabilities of a Conventional SEM Using an Off-the-Shelf Transmission Detector."
Paper presented at Microscopy & Microanalysis 2015 Meeting, Portland, OR, United States. August 2, 2015 - August 6, 2015.

be assigned to the Au particles. Figure 2c shows that Au particles hidden under an agglomeration of TiO_2 particles are not visible when viewed with a conventional secondary electron detector, but are easily discernible when imaging with the transmission detector and annular aperture system.

This work shows that a relatively low cost upgrade can significantly extend the analytical capabilities of practically any SEM [8].

References:
[1] T. Klein, *et al*., in Advances in Imaging and Electron Physics, P.W. Hawkes (Ed.), **171** (2012), 297.
[2] N. Hondow *et al*. Nanotoxicology **5** (2011) 215.
[3] M. Kuwajima *et al*. PLOS ONE **8** (2013) e59573.
[5] S. D. Findlay *et al*. Ultramicroscopy **110** (2010) 903.
[6] G. Rubin et al. Ultramicroscopy **113** (2012) 131.
[7] J. M. Cowley, J. Electron Micros. **50** (2001) 147.
[8] J. Holm acknowledges funding from the National Research Council Postdoctoral Research Associateship Program.

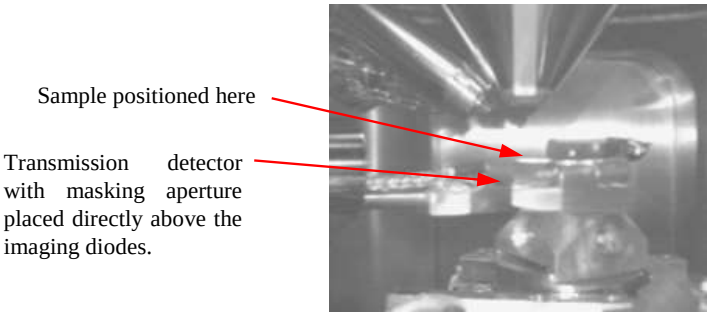
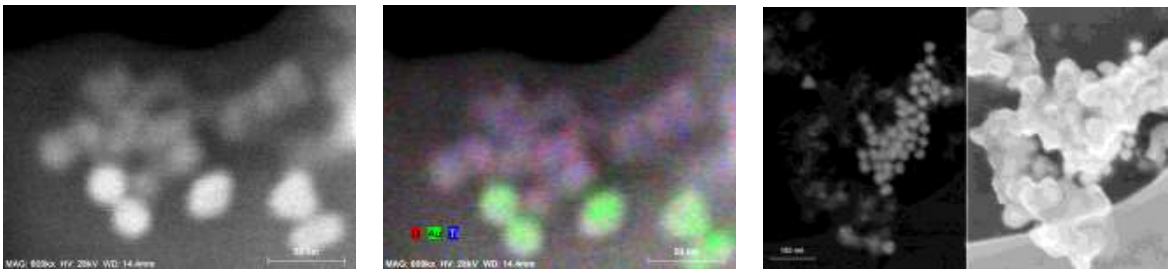


Figure 1. A view of the SEM chamber interior with the transmission detector, sample holder, and EDS detector inserted. This is the setup used to generate the images shown in Fig. 2. Because of the camera angle, the sample and aperture are not visible in this image.



(a) (b) (c)

Figure 2. HAADF transmission images showing 30 nm Au nanoparticles and ≈ 30 nm TiO_2 nanoparticles on a carbon substrate. The detector gain has been adjusted to show the carbon substrate. (a) An image showing intensity differences between the Au, TiO_2, and C. (b) An X-ray map superimposed on the HAADF image. (c) Split-screen image showing Au particles under a pile of TiO_2 particles. (HAADF image is on the left, secondary electron image is on the right.)

Holm, Jason; Keller, Robert.
"Analytical Transmission Scanning Electron Microscopy: Extending the Capabilities of a Conventional SEM Using an Off-the-Shelf Transmission Detector."
Paper presented at Microscopy & Microanalysis 2015 Meeting, Portland, OR, United States. August 2, 2015 - August 6, 2015.

Strategies for Parallel Markup

Bruce R. Miller

Applied and Computational Mathematics Division
National Institute of Standards and Technology, Gaithersburg, MD, USA

Abstract. Cross-referenced parallel markup for mathematics allows the combination of both presentation and content representations while associating the components of each. Interesting applications are enabled by such arrangements: interaction with parts of the presentation to manipulate and query the corresponding content; enhanced search indexing. Although the idea of such markup is hardly new, effective techniques for creating and manipulating it are more difficult than it appears. Since the structures and tokens in the two formats often do not correspond one-to-one, decisions and heuristics must be developed to determine in which way each component refers to and is referred to by components of the other representation. Conversion between fine and coarse-grained parallel markup complicates XML identifier (ID) assignments. In this paper, we will describe the techniques developed for LaTeXML, a TeX/LaTeX to XML converter, to create cross-referenced parallel MathML. While not yet considering LaTeXML's content MathML to be truly useful, the current effort is a step towards that continuing goal.

1 Introduction

Parallel markup for mathematics provides the capability of providing alternative representations of the mathematical expression, in particular, both the presentation form of the mathematics, i.e. its appearance, along with the content form, i.e. its meaning or semantics. Cross-linking between the two forms provides the connection between them such that one can determine the meaning associated with every visible fragment of the presentation and, conversely, the visible manifestation of each semantic sub-expression. Thus cross-linked parallel markup provides not only the benefits of the presentation and content forms, individually, but support many other applications such as: hybrid search where both the presentation and content can be taken into account simultaneously; interactive applications where the visual representation forms part of the user-interface, but supports computations based on the content representation.

Of course, the *idea* of parallel markup is hardly new. The m:semantics element has been part of the MathML specification [1] since the first version, in 1998! What seems to be missing are effective strategies for creating, manipulating and using this markup. Fine-grained parallelism is when the smallest sub-expressions are represented in multiple forms, whereas with coarse-grained parallelism the entire expression appears in several forms. Fine-grained parallelism is generally

Miller, Bruce.
"Strategies for Parallel Markup."
Paper presented at International Conference on Intelligent Computer Mathematics, CICM 2015, Washington, DC, United States. July 13, 2015 - July 17, 2015.

easier to create initially, and particularly when one deals with complex 'transfix' notations, or wants to preserve the appearance, but can infer the semantic intent of each sub-expression. Coarse-grained is often required by applications which may understand only a single format, or are unable to disentangle the fined-grained structure. HTML5 [3] only just barely accepts coarse-grained parallel markup, for example, by ignoring all but the first branch. Conversion from fine to coarse-grained is not inherently difficult, it can be carried out by a suitable walk of the expression tree for each format. But what isn't so clear is how to maintain the associations between the symbols and structures in the two trees. Indeed, there is typically no one-to-one correspondence between the elements of each format. Fine-grained parallelism, by itself, doesn't guarantee a clear association between all the symbols between the branches.

Our context here is LaTeXML, a converter from TeX/LaTeX to XML, and thence to web appropriate formats such as HTML, MathML and OpenMath. Input documents range from highly semantic markup such as sTeX [4], to intermediate such as used in the Digital Library of Mathematical Functions (DLMF) [6], to fairly undisciplined, purely presentational, markup as found on arXiv [8]. TeX induces high expectations for quality formatting forcing us to preserve the presentation of math. Meanwhile, the promise of global digital mathematics libraries and the potential reuse of a legacy of mathematics material encourages us to push as far as possible the extraction of content from such documents. At the very least, we should preserve whatever semantics is available in order to enable other technologies and research, such as LLaMaPuN [2], to resolve the remaining ambiguities. Getting ahead of ourselves, our first 'mini' strategy is just that: create fine-grained parallel markup at the first opportunity that presentation and semantics are available, in order to preserve them both.

In this paper, we describe the markup used in LaTeXML both for macros with known semantics, and for the result of parsing, and strategies for conversion to cross-linked, parallel markup combining Presentation MathML (pMML) and Content MathML (cMML). It should be noted that this does not mean that LaTeXML is producing useful quality cMML; the current work is a stepping stone towards that long-term goal.

Even in situations where only presentation markup is used, such as (currently) DLMF[1] where most symbols have always been hyper-linked to their definitions and the presentation-based search indexing is enhanced by the corresponding semantic content [7], the proposed strategies create a proper association between presentation and content resulting in a much cleaner implementation than the ad-hoc methods previously employed. Moreover, it is more complete, allowing the linkage to definitions to be extended to less textual operators such as binomials, floor, 3-j symbols, etc.

Miller, Bruce.
"Strategies for Parallel Markup."
Paper presented at International Conference on Intelligent Computer Mathematics, CICM 2015, Washington, DC, United States. July 13, 2015 - July 17, 2015.

Listing 1.1. Internal representation of $a + F(a, b)$, after parsing to XMath (assuming F as a function)

```
<XMApp>
  <XMTok meaning="plus" role="ADDOP">+</XMTok>
  <XMTok role="ID" font="italic">a</XMTok>
  <XMDual>
    <XMApp>
      <XMRef idref="m1.1" />
      <XMRef idref="m1.2" />
      <XMRef idref="m1.3" />
    </XMApp>
    <XMApp>
      <XMTok role="FUNCTION" xml:id="m1.1" font="italic">F</XMTok>
      <XMWrap>
        <XMTok role="OPEN" stretchy="false">(</XMTok>
        <XMTok role="ID" xml:id="m1.2" font="italic">a</XMTok>
        <XMTok role="PUNCT">,</XMTok>
        <XMTok role="ID" xml:id="m1.3" font="italic">b</XMTok>
        <XMTok role="CLOSE" stretchy="false">)</XMTok>
      </XMWrap>
    </XMApp>
  </XMDual>
</XMApp>
```

2 Motivation

Before diving into examples, a brief introduction to LaTeXML's internal mathematics markup, informally called XMath, is in order. This markup, inspired by OpenMath and both pMML and cMML, is intentionally hybrid in order to capture both the presentation and content properties of the mathematical objects throughout the step-wise processing from raw TeX markup, through parsing and, ultimately, semantic annotation. The main elements of concern are:

XMApp generalized application (think m:apply or om:OMA);
XMTok generalized token (think m:mi, m:mo, m:mn, m:csymbol);
XMDual parallel markup container of the content and presentation branches;
XMRef shares nodes between branches of XMDual, via xml:id and idref attributes;
XMWrap container of unparsed sequences of tokens or subtrees (think m:mrow).

(Please see the online manual[2] for more details.)

By way of motivation, consider the simple example in Listing 1.1. The role attribute on tokens indicates the syntactic role that it plays in the grammar; in this case, we've asserted that F is a function, allowing the expression to be parsed. At the top-level, the sum requires no special parallel treatment since the presentation for infix operators is trivially derived from the content form (i.e. the application of '+' to its arguments). The application of F to its arguments benefits somewhat from parallel markup. This is a typical situation with the fine-grained XMDual: the content branch is the application of some function or

[1] http://dlmf.nist.gov/
[2] http://dlmf.nist.gov/LaTeXML/manual/

Miller, Bruce.
"Strategies for Parallel Markup."
Paper presented at International Conference on Intelligent Computer Mathematics, CICM 2015, Washington, DC, United States. July 13, 2015 - July 17, 2015.

Listing 1.2. MathML representation of $a + F(a, b)$

```
<math display="block" alttext="a+F(a,b)" class="ltx_Math" id="m1">
  <semantics id="m1a">
    <mrow xref="m1.7.cmml" id="m1.7">
      <mi xref="m1.4.cmml" id="m1.4">a</mi>
      <mo xref="m1.5.cmml" id="m1.5">+</mo>
      <mrow xref="m1.6.cmml" id="m1.6d">
        <mi xref="m1.1.cmml" id="m1.1">F</mi>
        <mo xref="m1.6.cmml" id="m1.6e">&ApplyFunction;</mo>
        <mrow xref="m1.6.cmml" id="m1.6c">
          <mo xref="m1.6.cmml" id="m1.6" stretchy="false">(</mo>
          <mi xref="m1.2.cmml" id="m1.2">a</mi>
          <mo xref="m1.6.cmml" id="m1.6a">,</mo>
          <mi xref="m1.3.cmml" id="m1.3">b</mi>
          <mo xref="m1.6.cmml" id="m1.6b" stretchy="false">)</mo>
        </mrow>
      </mrow>
    </mrow>
    <annotation-xml id="m1b" encoding="MathML-Content">
      <apply xref="m1.7" id="m1.7.cmml">
        <plus xref="m1.5" id="m1.5.cmml"/>
        <ci xref="m1.4" id="m1.4.cmml">a</ci>
        <apply xref="m1.6d" id="m1.6.cmml">
          <ci xref="m1.1" id="m1.1.cmml">F</ci>
          <ci xref="m1.2" id="m1.2.cmml">a</ci>
          <ci xref="m1.3" id="m1.3.cmml">b</ci>
        </apply>
      </apply>
    </annotation-xml>
    <annotation id="m1c" encoding="application/x-tex">a+F(a,b)</annotation>
  </semantics>
</math>
```

operator (here F) to arguments (here a, b), but they are represented indirectly using XMRef to point to the corresponding sub-expressions within the presentation. While one could represent the delimiters and punctuation as attributes (as in MathML's m:mfenced), that loses attributes of those attributes such as stretchiness, size or even color. A more compelling case is made by complex transfix notations or semantic macros, as we will shortly see.

However, this simple example already hints at a hidden complexity. Converting to either pMML and cMML is straightforward (given rules for mapping XMath elements to MathML): simply walk the tree, depth-first, following each XMRef to the referenced node and choosing the first or second branch of XMDual for content or presentation, respectively. Even cross-linking is straightforward in the absence of XMDual, when the generated content or presentation nodes are 'sourced' from the same XMath node (F, a, and b, in the example): simply assign ID's to the source XMath node and the generated nodes; record the association between them; afterwards, the presentation and content nodes that were sourced from the same ID are connected by getting an xref attribute referring each to the other. But with XMDual one has not only to determine when the generated nodes are related, one has to contend with extra tokens; in the example, the parentheses and comma appear only in the presentation. Presumably, those tokens

should be associated with the *application* of F, as would the containing m:mrow. The desired result is shown in Listing 1.2.

A fuller illustration of the issues encountered in typical LaTeX markup combines complex transfix notations and semantic macros, such as:

```
\left\langle\Psi\middle|\mathcal{H}\middle|\Phi\right\rangle
 +  \defint{a}{b}{F(x)}{x}
```

This example, whose internal form is shown in Listing 1.3, involves quantum-mechanics notations, which LaTeXML's parser is happily able to recognize. Additionally, we've introduced a semantic macro `\defint` to represent definite integration, which will be transformed to so-called 'Pragmatic' Content MathML form, to enhance the illustration with a many-to-many correspondence. (The implementation of `\defint` is not difficult, but outside the scope of this article)

3 Main Strategies

We will see that a key part of the method is determining which nodes are visible to presentation, to content or to both branches. This is easily determined by an algorithm like mark-and-sweep garbage collection. Simply traverse the tree from the root following the first branch of each XMDual to mark all nodes as visible to presentation, and repeat for the second branch to mark content visibility. The analogy is apt, as this also allows pruning of nodes that are not visible at all, as sometimes occurs with complex macro usage — another mini-strategy.

A few definitions will also be relevant. During the depth-first tree traversal transforming XMath into either pMML or cMML, the *current node* is the XMath node being transformed. The *current container* is the XMDual node (if any) containing the current node; in the context of this discussion, an XMath container is always the application of some function or operator. We'll call the latter the *current operator*.

We will ascribe to each generated target node (pMML or cMML) an XMath *source node* that can be considered 'responsible' for the target node. In the common, simplest case, a current node that is visible to both branches and generates a token node in the target can be used as the source node.

A special case occurs when the target MathML element is a container; these generally do not correspond to symbols, and ought to be associated with the nearest application (think m:apply or m:mrow)[3]. In this case, the source should be the nearest ancestor XMDual of the current node, that is the *current container*.

Similar reasoning applies when a token (non-container) symbol is generated from an XMath token which is not visible to the opposite branch; typically the notational icing of some transfix. We presume that is the only visible manifestation of the *current operator*, and so that is taken as the source node.

[3] Exceptions are m:msqrt or m:menclose where they tend to represent *both* the application of an operation and yet are the only visible manifestation of the operator! However, we also note that a common use of cross-linking in HTML is to turn them into href links; but HTML does not allow nested links!

Listing 1.3. Internal representation of $\langle\Psi|\mathcal{H}|\Phi\rangle + \int_a^b F(x)dx$ as XMath

```
<XMApp>
  <XMTok meaning="plus" role="ADDOP">+</XMTok>
  <XMDual>
    <XMApp>
      <XMTok meaning="quantum-operator-product" />
      <XMRef idref="m2.5" />
      <XMRef idref="m2.6" />
      <XMRef idref="m2.7" />
    </XMApp>
    <XMWrap>
      <XMTok role="OPEN">⟨</XMTok>
      <XMTok role="ID" xml:id="m2.5">Ψ</XMTok>
      <XMTok role="CLOSE" stretchy="true">|</XMTok>
      <XMTok role="ID" xml:id="m2.6" font="caligraphic">H</XMTok>
      <XMTok role="OPEN" stretchy="true">|</XMTok>
      <XMTok role="ID" xml:id="m2.7">Φ</XMTok>
      <XMTok role="CLOSE">⟩</XMTok>
    </XMWrap>
  </XMDual>
  <XMDual>
    <XMApp>
      <XMTok meaning="hack-definite-integral" role="UNKNOWN" />
      <XMRef idref="m2.1" />
      <XMRef idref="m2.2" />
      <XMRef idref="m2.3" />
      <XMRef idref="m2.4" />
    </XMApp>
    <XMApp>
      <XMApp>
        <XMTok role="SUPERSCRIPTOP" scriptpos="post2" />
        <XMApp>
          <XMTok role="SUBSCRIPTOP" scriptpos="post2" />
          <XMTok mathstyle="display" meaning="integral" role="INTOP">∫</XMTok>
          <XMTok role="ID" xml:id="m2.1" font="italic">a</XMTok>
        </XMApp>
        <XMTok role="ID" xml:id="m2.2" font="italic">b</XMTok>
      </XMApp>
      <XMApp>
        <XMTok meaning="times" role="MULOP">⁢</XMTok>
        <XMDual xml:id="m2.3">
          <XMApp>
            <XMRef idref="m2.3.1" />
            <XMRef idref="m2.3.2" />
          </XMApp>
          <XMApp>
            <XMTok role="FUNCTION" xml:id="m2.3.1" font="italic">F</XMTok>
            <XMWrap>
              <XMTok role="OPEN" stretchy="false">(</XMTok>
              <XMTok role="UNKNOWN" xml:id="m2.3.2" font="italic">x</XMTok>
              <XMTok role="CLOSE" stretchy="false">)</XMTok>
            </XMWrap>
          </XMApp>
        </XMDual>
        <XMApp>
          <XMTok meaning="differential-d" role="DIFFOP" font="italic">d</XMTok>
          <XMTok role="UNKNOWN" xml:id="m2.4" font="italic">x</XMTok>
        </XMApp>
      </XMApp>
    </XMApp>
  </XMDual>
</XMApp>
```

In the example, the angle brackets and vertical bars are thus ascribed to the `quantum-operator-product` operator.

In pseudo-code, the *source* node for a given *target* is thus:

```
if target is a container
  if current container exists
    current container
  else
    current
else if target is visible in both branches
  current
else if current container exists
  if current operator is hidden from presentation
    current operator
  else
    current container
else
  current
```

Once this ascription of source nodes is done, the cross-referencing between the generated targets is easily established: the xref of a pMML (cMML) node is the cMML (pMML, respectively) node that was ascribed to the same source XMath node; if multiple nodes were ascribed to that source node, the first target node, in document order, is the sensible choice. Applying this method to the example from Listing 1.3 yields Listing 1.4, where we can see, for example, that the angle brackets and vertical bars are associated with the `quantum-operator-product` operator while the various m:bvar, m:lowlimit, etc, are properly associated with the integral, *not* the integral operator.

4 Outlook

We have described a set of strategies for generating parallel markup with cross-references consisting of encouraging fine-grained parallel structures, mark-and-sweep to detect nodes visible to presentation or content and to garbage-collect, and a method to determine related nodes within each branch of the parallel markup.

Parallel markup must also be adapted to larger structures such as eqnarray, and AMS alignments with intertext containing multiple formula and/or document-level text markup. While the fundamental issue is the same — separating presentation and content forms — this seems to demand a distributed markup that separates the presentation and content forms into distinct math containers. LaTeXML currently has an ad-hoc, but not entirely satisfactory solution for this, but we will experiment with adapting the methods described here. However, it remains to be seen whether cross-referencing across separate math containers can be made useful.

And, now that generating Content MathML is more fun, we must continue working towards generating *good* Content MathML. Ongoing work will attempt

Listing 1.4. MathML representation of $\langle\Psi|\mathcal{H}|\Phi\rangle + \int_a^b F(x)dx$

```
<math display="block" alttext="..." class="ltx_Math" id="m2">
  <semantics id="m2a">
    <mrow xref="m2.13.cmml" id="m2.13">
      <mrow xref="m2.9.cmml" id="m2.9">
        <mo xref="m2.8.cmml" id="m2.8">&LeftAngleBracket;</mo>
        <mi mathvariant="normal" xref="m2.5.cmml" id="m2.5">&Psi;</mi>
        <mo xref="m2.8.cmml" id="m2.8a" stretchy="true" fence="true">|</mo>
        <mi xref="m2.6.cmml" id="m2.6" class="ltx_font_mathcaligraphic">&HilbertSpace;</mi>
        <mo xref="m2.8.cmml" id="m2.8b" stretchy="true" fence="true">|</mo>
        <mi mathvariant="normal" xref="m2.7.cmml" id="m2.7">&Phi;</mi>
        <mo xref="m2.8.cmml" id="m2.8c">&RightAngleBracket;</mo>
      </mrow>
      <mo xref="m2.10.cmml" id="m2.10">+</mo>
      <mrow xref="m2.12.cmml" id="m2.12c">
        <msubsup xref="m2.12.cmml" id="m2.12">
          <mo xref="m2.11.cmml" id="m2.11" symmetric="true" largeop="true">&int;</mo>
          <mi xref="m2.1.cmml" id="m2.1">a</mi>
          <mi xref="m2.2.cmml" id="m2.2">b</mi>
        </msubsup>
        <mrow xref="m2.12.cmml" id="m2.12b">
          <mrow xref="m2.3.cmml" id="m2.3c">
            <mi xref="m2.3.1.cmml" id="m2.3.1">F</mi>
            <mo xref="m2.3.cmml" id="m2.3d">&ApplyFunction;</mo>
            <mrow xref="m2.3.cmml" id="m2.3b">
              <mo xref="m2.3.cmml" id="m2.3" stretchy="false">(</mo>
              <mi xref="m2.3.2.cmml" id="m2.3.2">x</mi>
              <mo xref="m2.3.cmml" id="m2.3a" stretchy="false">)</mo>
            </mrow>
          </mrow>
          <mo xref="m2.11.cmml" id="m2.11a">&InvisibleTimes;</mo>
          <mrow xref="m2.12.cmml" id="m2.12a">
            <mo xref="m2.11.cmml" id="m2.11b">d</mo>
            <mi xref="m2.4.cmml" id="m2.4">x</mi>
          </mrow>
        </mrow>
      </mrow>
    </mrow>
    <annotation-xml id="m2b" encoding="MathML-Content">
      <apply xref="m2.13" id="m2.13.cmml">
        <plus xref="m2.10" id="m2.10.cmml"/>
        <apply xref="m2.9" id="m2.9.cmml">
          <csymbol xref="m2.8" id="m2.8.cmml" cd="latexml">quantum-operator-product</csymbol>
          <ci xref="m2.5" id="m2.5.cmml">normal-&Psi;</ci>
          <ci xref="m2.6" id="m2.6.cmml">&HilbertSpace;</ci>
          <ci xref="m2.7" id="m2.7.cmml">normal-&Phi;</ci>
        </apply>
        <apply xref="m2.12c" id="m2.12.cmml">
          <int xref="m2.11" id="m2.11.cmml"/>
          <bvar xref="m2.12c" id="m2.12a.cmml">
            <ci xref="m2.4" id="m2.4.cmml">x</ci>
          </bvar>
          <lowlimit xref="m2.12c" id="m2.12b.cmml">
            <ci xref="m2.1" id="m2.1.cmml">a</ci>
          </lowlimit>
          <uplimit xref="m2.12c" id="m2.12c.cmml">
            <ci xref="m2.2" id="m2.2.cmml">b</ci>
          </uplimit>
          <apply xref="m2.3c" id="m2.3.cmml">
            <ci xref="m2.3.1" id="m2.3.1.cmml">F</ci>
            <ci xref="m2.3.2" id="m2.3.2.cmml">x</ci>
          </apply>
        </apply>
      </apply>
    </annotation-xml>
    <annotation id="m2c" encoding="application/x-tex">...</annotation>
  </semantics>
</math>
```

to establish appropriate OpenMath Content Dictionaries, probably in a FlexiFormal sense [5], improved math grammar, and exploring semantic analysis.

References

1. Ausbrooks, R., Buswell, S., Carlisle, D., Chavchanidze, G., Dalmas, S., Devitt, S., Diaz, A., Dooley, S., Hunter, R., Ion, P., Kohlhase, M., Lazrek, A., Libbrecht, P., Miller, B., Miner, R., Sargent, M., Smith, B., Soiffer, N., Sutor, R., Watt, S.: Mathematical Markup Language (MathML) version 3.0. W3C Recommendation, World Wide Web Consortium (W3C) (2010), `http://www.w3.org/TR/MathML3`
2. Ginev, D., Jucovschi, C., Anca, S., Grigore, M., David, C., Kohlhase, M.: An architecture for linguistic and semantic analysis on the arXMLiv corpus. In: Applications of Semantic Technologies (AST) Workshop at Informatik 2009 (2009).
3. Hickson, I., Berjon, R., Faulkner, S., Leithead, T., Navara, E.D., O'Connor, E., Pfeiffer, S.: HTML5. W3C Recommentation, World Wide Web Consortium (W3C) (2014), `http://www.w3.org/TR/html5/`
4. Kohlhase, M.: Using LaTeX as a semantic markup format. Mathematics in Computer Science 2(2), 279–304 (2008).
5. Kohlhase, M.: The flexiformalist manifesto. In: Voronkov, A., Negru, V., Ida, T., Jebelean, T., Petcu, D., ane Daniela Zaharie, S.M.W. (eds.) 14th International Workshop on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2012). pp. 30–36. IEEE Press, Timisoara, Romania (2013).
6. Miller, B.R., Youssef, A.: Technical aspects of the digital library of mathematical functions. Annals of Mathematics and Artificial Intelligence 38(1-3), 121–136 (2003).
7. Miller, B.R., Youssef, A.: Augmenting presentation mathml for search. In: Proceedings of Conference on Intelligent Computer Mathematics. Lecture Notes in Artificial Intelligence, vol. 5144. Springer, Heidelberg (2008).
8. Stamerjohanns, H., Kohlhase, M., Ginev, D., David, C., Miller, B.: Transforming large collections of scientific publications to XML. Mathematics in Computer Science 3(3), 299–307 (2010).

Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections

Ossama Younis and Nader Moayeri
National Institute of Standards and Technology
Gaithersburg, MD 20899, USA

Abstract— **Traffic control at road intersections is based on traffic lights (TLs). The control mechanism typically used for traffic lights operates based on a periodic schedule to change the light (red/yellow/green). In many cases, a different schedule is used in late night/early morning hours. This fixed control mechanism does not adequately account for changing traffic conditions, and is unaware of/unresponsive to congestion. In this work, we propose a framework for dynamic traffic light control at intersections. The framework relies on a simple sensor network to collect traffic data and includes novel protocols for traffic flow control to handle congestion and facilitate flow. We show that our proposed algorithms have low overhead and are practical to employ in live traffic flow scenarios. Through extensive simulations, we demonstrate the benefits of our framework in optimizing traffic flow metrics, such as traffic throughput, average vehicle waiting time, and vehicle waiting line length.**

Keywords-Traffic light control; traffic flow optimization; distributed algorithms; sensor networks.

I. INTRODUCTION

Cyber-Physical Systems (CPS) have been recently proposed to define how the computing world interacts with and manipulates the physical world. Researchers are developing CPS for application domains such as environmental monitoring, smart homes, and smart manufacturing. While the CPS idea is rising, several questions require answers. For example, it is not clear if the computing foundations are adequate for supporting and using CPS [1]. Some argue that CPS requires top-to-bottom rethinking of computation.

Improving traffic flow and handling congestion are classic problems in transportation systems. Research in Civil Engineering has focused on improving the road conditions, adding/switching traffic lanes, and controlling traffic lights to improve traffic flow particularly on main/congested roads. New research focuses on utilizing CPS to optimize traffic flow and improve the driver's experience on the road. Most previous research has shown the "benefits" of using controlled traffic lights. However, practical protocols are still needed to control lights and optimize traffic flow.

In this work, we present a framework for optimizing traffic flow using dynamic traffic lights. Our framework includes collecting road conditions by sensors deployed in road side units (RSUs) and using algorithms to decide on when to change traffic lights to alleviate congestion. Our proposed techniques are fast, simple, low-overhead, and thus can be easily deployed. Through analysis and simulations, we demonstrate the benefits of the proposed framework and protocols.

The rest of this paper is organized as follows. Section II describes the system/road model and explores the possible design approaches for traffic light control. Sections III and IV present our solutions for traffic flow enhancement. Section V analyzes the proposed approaches: describes the overhead (or complexity), outlines the possible triggers to change the traffic lights ahead of their typical schedule, and discusses how our proposed approaches extend to global traffic light control (i.e., in a region or town). Section VI evaluates the proposed algorithms in light/heavy traffic scenarios. Section VII overviews related research on traffic flow control. Finally, Section VIII concludes this work.

II. SYSTEM MODEL AND DESIGN APPROACHES FOR DYNAMIC TRAFFIC LIGHT CONTROL

A. *Road/Traffic Model*

We assume the following about the traffic pattern, the vehicle capabilities, and the road intersection:

- **Road:** Typical model with intersecting roads. A road has two opposite segments, each allowing traffic to go along one or more lanes.
- **Traffic lights:** Installed at the head of each road segment. The traffic lights at two opposite/intersecting segments always show the same/opposite colors.
- **Traffic model:** Vehicles arrive according to an arbitrary stochastic model. We propose traffic light control techniques that are independent of the arrival model. We assume that all vehicles at the light want to proceed forward, and leave left/right turns for future work.
- **Vehicle detection:** By exploiting either RSU sensors or vehicle-installed devices, detect/report the vehicles that arrive at the intersection on each of the four segments.

B. *Objectives*

Develop protocol(s) to control/optimize the traffic flow. These protocols collect traffic flow information and utilize it to optimize traffic flow metrics, such as throughput, delay, or waiting line length. The protocols should have low communication overhead and fast decision-making to be practically applicable on the road. The cost of the proposed protocols should also be minimal to motivate adoption and deployment. This is why it is desirable to propose solutions for deployment and operation. Such solutions may trade one performance metric for another.

Younis, Ossama; Moayeri, Nader.
"Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections."
Paper presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar. April 3, 2016 - April 6, 2016.

C. *Design Approaches*

We categorize the design of dynamic traffic control mechanisms into two possible approaches:

Design Approach 1: Deploy sensor infrastructure in RSUs to collect information about the road conditions (e.g., the number of vehicles waiting for a green light). Also deploy a traffic light controller (TLC) at the traffic light to control its state based on an algorithm that uses information from the RSUs. Alternatively, the infrastructure can be sensors placed at road lanes. This design option assumes that the sensors are able to communicate their data to the TLC. We name this approach Road-based Infrastructure Traffic-light Control (**RITCO**).

Design Approach 2: Use distributed infrastructure, e.g., (1) a traffic light controller (TLC) at the traffic light, and (2) vehicle communication devices to report the vehicles' arrivals and locations (road segments) to the control unit (CU). This reduces the amount of road infrastructure, but imposes requirements on the vehicles to support the TLC. We name this design approach Vehicle-based Infrastructure Traffic-light Control (**VITCO**).

D. *Which Design Approach to Select*

Several issues need to be considered in order to adopt a particular design:

- **Complexity/Cost:** Evaluate processing/communication complexity, in addition to cost. Designs with significant processing can be acceptable if: (1) the computing devices are expected to be capable of rapidly carrying out the computations, and (2) the processing complexity may reduce the communication complexity.
- **Applicability:** Describes which mechanism is more practical in the transportation road network model, and therefore, is more applicable. This involves studying effectiveness, ease of deployment, and cost.
- **Performance metrics:** The metrics are defined in Section VI.A (e.g., traffic throughput/waiting time). The performance of the proposed technique(s) is a major factor in deciding which design approach is more suitable for the traffic model that we optimize.

III. ROAD-BASED INFRASTRUCTURE TRAFFIC-LIGHT CONTROL (RITCO)

The RITCO protocol employs "Design approach 1" in Section II.C. It requires:

- **Traffic light controller (TLC):** (1) Co-located with the traffic light and (2) collects information about the current road conditions from RSUs and uses it to make light control decisions.
- **Sensors at RSUs:** Each road segment is equipped with an RSU installed several meters away from the intersection to detect vehicle arrivals. Assuming that the traffic light is going to change to red at time T_{Red}, each of the two RSUs installed on segments with the red light is notified of the impending change at time (T_{Red}-Δt), where Δt is a small positive number (interval), to start counting the vehicles that will be waiting at the TL. The counted number of vehicles is periodically reported to the TLC. Note that if multiple sensors are used in each segment, then coordination is needed to avoid counting a vehicle multiple times. Using a single sensor eliminates the communication overhead. However, one sensor may not provide sufficient coverage to detect all vehicles.

A. *The RITCO Protocol*

Table 1 lists the pseudo-code of the RITCO protocol. $T_{curr} \equiv$ current time, $\Delta t \equiv$ very small time interval, $T_{red}/T_{yellow}/T_{green} \equiv$ Red/Yellow/Green light intervals, $T_{change} \equiv$ time until light changes to red/green, numVehicles ≡ number of vehicles, VID ≡ Vehicle ID (a unique ID, e.g., license plate number).

Table 1. The RITCO Protocol Details

Steps at TLC (Current Light Green→Change to Red):

- **Initialize:** numVehicles←0, switchLightFlag←OFF
- TLC announces impending TL change to red to RSUs.
- **While** (($T_{curr}+\Delta t \geq T_{red}$) && (switchLightFlag is OFF))
  - Collect sensor (RSU) readings
  - If (m new vehicles are detected at RSU)
    - numVehicles ← numVehicles+m
    - Save VID's to vehicle list
  - UpdateFlag(switchLightFlag)
- **End While**
- **If** (switchLightFlag is ON)
  - Perform TL_Change (Red→Green)
- The TLC announces impending TL change to RSUs.

Steps at Sensor (RSU on Pole or Road Lane):

- **Initialize:** Vehicle counter (numVehicles ← 0).
- RSU prepares to activate the sensing process:
  - While ($T_{curr}+\Delta t \leq T_{change}$)
    - Wait/update T_{curr} to the current time.
- Activate RSU for vehicle counting
- Whenever (RSU detects a vehicle)
  - numVehicles ← numVehicles+1
- Every k seconds, RSU checks:
  - If (numVehicles > 0)
    - Report numVehicles to the TLC
    - RSU resets its numVehicles to 0.
- Based on the upcoming TL change to green, RSU decides when to stop counting & when to report its count to the TLC.

TL Change Procedure (Green → Red):

- Change green light to yellow.
- Wait for an interval T_{yellow} time.
- Change yellow light to red.

TL Change Procedure (Red→Green):

- Change red light to green.

UpdateFlag(flag):

- If (light switch trigger is satisfied)
  flag ← ON
- Else
  flag ← OFF

Younis, Ossama; Moayeri, Nader.
"Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections."
Paper presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar. April 3, 2016 - April 6, 2016.

IV. VEHICLE-BASED INFRASTRUCTURE TRAFFIC-LIGHT CONTROL (VITCO)

The VITCO protocol employs "Design Approach 2" in Section II.C. It requires:

- **Traffic light controller (TLC):** (1) Co-located with the traffic light, and (2) interprets/executes the decisions made by the VITCO protocol for light control.
- **Vehicle Radio:** A vehicle device that is used to announce vehicle's arrival at the traffic light. The radio executes what was being done by the RSUs in RITCO, thus participating in the coordination process.

A. The VITCO Protocol

Table 2 provides the pseudo-code of the VITCO protocol (executed at every road segment). Symbol definitions are similar to those in Section III.A.

Table 2. The VITCO Protocol Details

VITCO: Elect a Coordinating Vehicle (Elect_CoVe):

- A vehicle V_1 arrives/stops at the TL.
- V_1 transmits a message for CoVe and neighbor discovery and sets an expiration timer T_1.
- If a response message announcing CoVe ID arrives (from a neighbor), cancel T_1 and find the path to that CoVe.
- If T_1 expires and no CoVe announcement has arrived, broadcast V_1 as a new CoVe.
- If later, V_1 receives a CoVe broadcast (after announcing itself as CoVe), arbitrate to pick the best CoVe. If selected, rebroadcast a CoVe announcement.
- If the other CoVe is better, then V_1 broadcasts a NON-CoVe message, including the other CoVe ID.

VITCO: Steps at CoVe (Lights Changing to Red):

- Vehicle V_i arrives at the traffic light.
- V_i discovers its neighbors and CoVe (Elect_CoVe).
- Assume the elected CoVe is V_c.
- numVehicles ← 0.
- While (TL_Timer has not expired)
  - If (V_i is V_c) Then
    - Broadcast a CoVe announcement
    - When (message_arrives_from a vehicle)
      - If (new vehicle)
      - Then Increment numVehicles
      - sendMsgToTLC(False, numVehicles)
      - EndIf
    - UpdateFlag(switchLightFlag)
    - If (switchLightFlag=ON) Then
      - SendMsgToTLC(True, numVehicles)
      - When the TLC confirms (TL changes)
      - Exit
    - EndIf
  - EndIf
  - Else // (V_i is not V_c)
    - V_i registers V_c as its CoVe
    - V_i discovers the communication path to V_c (if not single hop)
  - EndIf
- EndWhile

Note that the function Veh_SendMsgToTLC translates to function TLC_RecvMsg at the TLC (below).

TLC_RecvMsg(Bool changeTL, int numVehicles)

- If (changeTL = TRUE) Then
  - TLC.numVehicles = 0
  - Change the state of each traffic light
  - Exit Procedure
- EndIf
- If (numVehicles > 0)
  - TLC.numVehicles= numVehicles
  - If (TLC.numVehicles >= MAX_WAITING_VEH)
    - TLC.numVehicles = 0
    - If (TL is red)
      - Then TL_Change(Red → Green)
      - Else TL_Change(Green → Red)
    - EndIf
    - Exit Procedure
  - End If
- EndIf

V. PROTOCOL ANALYSIS

A. Properties/Overhead

(1) Communication: RITCO is independent of the vehicles on the road, i.e., requires no communication between the vehicles and the infrastructure. VITCO has more communication cost.

(2) Processing: RITCO assigns the work to road infrastructure (RSUs or devices attached to the traffic light). VITCO, on the other hand, puts most of the processing work on the Coordinating Vehicle (CoVe), and communication work on all vehicles.

(3) Cost: RITCO puts all the burden on the city/county (road infrastructure & traffic light control equipment). On the other hand, VITCO shares the cost between the city/county and the vehicle owners.

(4) Deployment: It is much easier/safer to rely on RITCO because of its independence of the vehicles on the road. It avoids the problems that may occur due to faulty vehicle equipment. This, however, requires much more cost than using VITCO. RITCO also avoids the transition period in which vehicles will need to deploy the necessary equipment to support VITCO.

(5) Security: RITCO is more secure than VITCO because it does not acquire information from the vehicles (just counts them autonomously). Decisions at VITCO are made by the vehicles, which presents vulnerability to greedy users (tampered control devices). To protect drivers' privacy, the vehicles could announce their presence using unique identifiers, such as a hash value of the vehicle identification number (VIN).

Younis, Ossama; Moayeri, Nader.
"Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections."
Paper presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar. April 3, 2016 - April 6, 2016.

(6) Complexity: Let N: the max number of vehicles waiting for the TL change, L: Number of lanes on the road. Thus, message exchange complexity/overhead per vehicle: O(N/L); messages to broadcast traffic light change: O(1); processing complexity: O(N) to register new vehicles. To decide whether/when to change TL: Can be as low as O(1) for heuristic-based algorithms.

B. Triggers to change traffic lights synchronously or asynchronously

The current traffic light transitions use fixed (pre-assigned) intervals for the different lights. Recently, in some areas, some forms of adaptive light transitions were employed. For example, prolonging the green light at main streets until a vehicle arrives at a side street. In this work, we consider a broader scope for early transitioning, and propose a detailed approach for adaptive lights. Two ways can change the TL (red→green or green→red):

- **Synchronously:** According to a timer expiration (which is being employed on the roads now).
- **Asynchronously:** In addition to waiting for timer expiration, the TL is changed if certain events trigger the need for light change.

Triggers for asynchronous change can be due to (but are not limited to) the occurrence of the following events:

1) The number of vehicles that are waiting for the green light exceeds a certain threshold. Note that such number of waiting cars can be considered on a single side of the TL or collectively on both sides of the TL. Considering the waiting vehicles on each side separately is more practical to avoid having a very long waiting line on any side.

2) No vehicles have crossed the intersection from the sides that have the green light for a certain time interval.

C. Regional/ global traffic light control

The framework (presented methods in Section III and Section IV) presents protocol algorithms for "local" traffic control at a single traffic light. Based on the reactive/independent nature of the proposed algorithms, the framework inherently addresses global traffic optimization. The reaction to congestion at a traffic light shifts the congestion problem toward other traffic lights, thus continuously trying to distribute the traffic load and relieve congested areas.

VI. EVALUATION OF PROPOSED APPROACHES

We evaluate the performance of the proposed Dynamic Traffic Light Control (DTLC) approach, in comparison to the legacy traffic light control (TLC). The evaluation does not distinguish between the two flavors of DTLC (RITCO and VITCO), because they differ only in the details of collecting traffic information but not in the TL control details. We first define the metrics that are used to evaluate our protocols, then we describe the experiments in detail.

A. Metrics/Parameters

We consider the following evaluation metrics: waiting time, waiting line length, and traffic throughput (definitions are given in Table 3):

Table 3. The metrics to evaluate DTLC

Metric	Definition
Average/Maximum waiting line length (Q_i)	At each side of the intersection, the average/maximum (number of vehicles per lane) among all lanes.
Average/Maximum waiting time (W_i)	The average/max time that a vehicle waits for change of the traffic light (red→green).
Throughput (T)	The number of vehicles that pass the intersection per unit time.

In our experiments, we vary two parameters: (1) number of arriving vehicles at the TL, and (2) number of TL cycles (multiple red-green transitions). The rest of the parameters are: #lanes=2, lane size=100 vehicle, vehicles arrival rate=90/lane, threshold of waiting vehicles=75, max# passing vehicles=150, red/green interval=20 sec, yellow interval=2 sec. Each result is the average of 10 random experiments.

B. Vary the Number of Arriving Vehicles

In the following experiments, we evaluate the different metrics when the number of arriving vehicles at the TL varies (the probability distribution of vehicle arrivals does not matter; only the number of arrivals matter). Figure 1 shows the throughput of the intersection point when the number of arriving vehicles varies from 20 to 100. The legacy TL system does not react to increasing traffic, while DTLC reacts to congestion by early switching the TL if the number of waiting vehicles exceed a threshold (e.g., 75). This is why significant throuhput gains are achieved by DTLC when the number of arriving vehicles exceeds such threshold. Figure 2 shows the number of waiting vehicles for green light. In DTLC, the waiting number of vehicles is upper-bounded by a limit, and early light switch is triggered when such limit is reached. Since this action is taken at both (intersecting) parts of the traffic light, this ensures that the waiting line never overflows and also that congestion is early handled. For the waiting time/vehicle, Figure 3 shows that vehicles wait slightly less when the DTLC approach is employed.

C. Vary the number of TL cycles (red-green interval)

We evaluate DTLC when the number of TL cycles varies (a cycle is a red light interval followed by a green light interval). The number of waiting vehicles is greater than 80. Figure 4 shows 250% traffic throughput improvement under heavy load. Figure 5 shows that the waiting line length is about 20% less with DTLC than the legacy TLC. Figure 6 illustrates that DTLC has less vehicle waiting time. This is intuitive, especially under heavy traffic that causes early light change.

D. Forced early TL change/Fairness

Experiments show that under high traffic load (i.e., exceeding the threshold), the TL change is forced at every cycle. Fairness is also achieved; i.e., the waiting time is similar for vehicles in corresponding locations from the traffic light.

Younis, Ossama; Moayeri, Nader.
"Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections."
Paper presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar. April 3, 2016 - April 6, 2016.

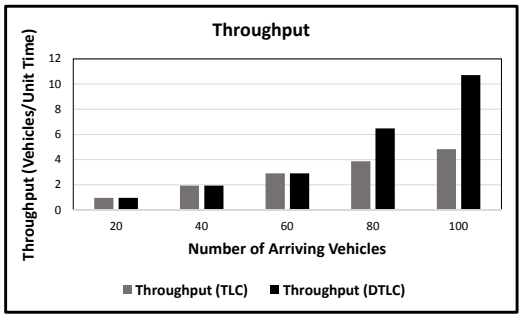


Figure 1. Throughput vs. Arriving Vehicles

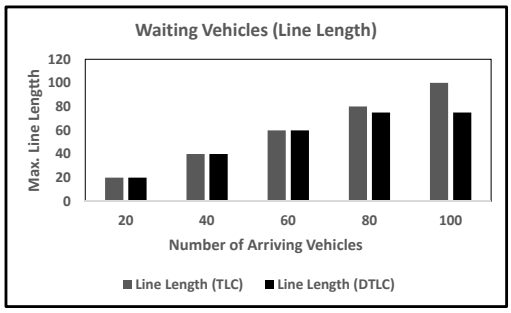


Figure 2. Waiting Line Length vs. Arriving Vehicles

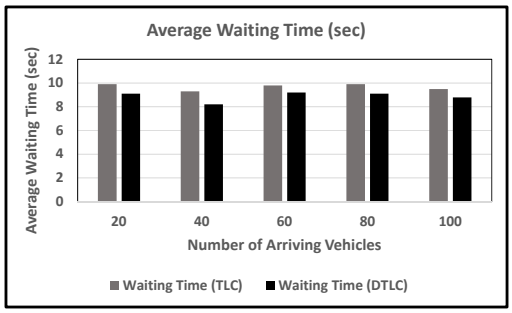


Figure 3. Waiting Time vs. Arriving Vehicles

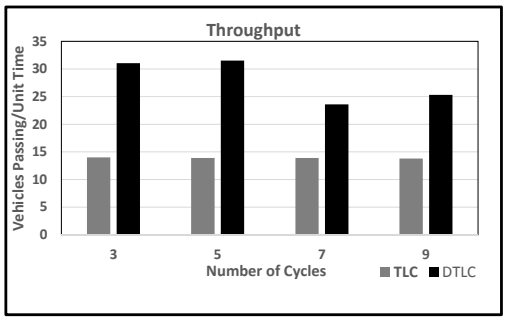


Figure 4. Throughput (Passing Vehicles/Unit Time) vs. the Number of Cycles

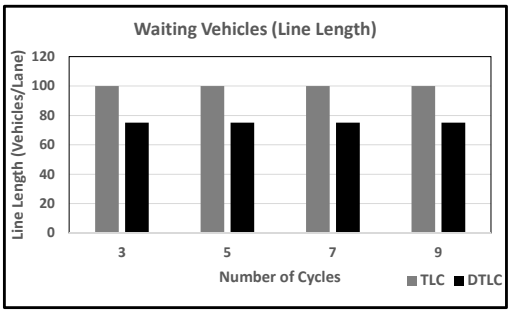


Figure 5. Waiting Line Length vs. the Number of Cycles

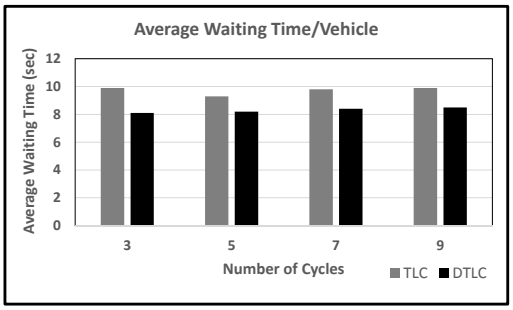


Figure 6. Waiting Time vs. the Number of Cycles

VII. RELATED WORK

The problem of how to control traffic lights to reduce congestion has been recently addressed in the literature. Two research approaches (Ferreira et al. [2] and Neudecker et al. [5]) are close to our work's objective and model, but differ in the solution. In [5], the authors addressed how to control traffic lights to facilitate better flow of traffic at an intersection. The vehicles on each road segment elect a leader, and all the leaders across the intersection compete to elect a virtual traffic light (VTL) leader. The technique, however, is not clear on how a leader is elected. It is important to design an approach that ensures that only one VTL leader is present at the intersection, and this VTL leader is the best candidate. Ferreira et al. [2] presented a virtual traffic light system. Their decentralized approach has two steps: (1) vehicles agree on a virtual traffic light leader at an intersection, and (2) the leader takes the role of a temporary virtual infrastructure and informs the neighboring vehicles of the traffic light schedule. When a leader leaves the intersection, it hands over the leadership to another vehicle. This approach is quite flexible, but it is not clear how the system ensures that a traffic light leader is elected (and how to ensure only one leader is elected). The work in [3] addresses the question: "to which extent can inter-vehicle communication improve wide area transportation network that consists of several cities?" Finally, [6] studies adaptive traffic lights based on car-to-car communication.

Several position papers have also presented the general problems of CPS and their applications. Tabuada [7] claims that a major research challenge is to understand how we can

Younis, Ossama; Moayeri, Nader.
"Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections."
Paper presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar. April 3, 2016 - April 6, 2016.

adapt the notion of locality induced by a network of embedded systems to the notion of locality. Abdelzaher [8] indicates that several trends impact the future of computer science, such as: (1) Moore's law, (2) widening human/machine bandwidth gap, and (3) high cost of lack of communication. These trends motivate the CPS development if challenges are overcome.

The CPS foundations are discussed in [9]. The claim is that the CPS problem is not the union of the cyber and physical problems, but rather the intersection. However, there are several limitations [10], including, (1) a critical gap between the emerging cyber-physical infrastructure and user environments, (2) existing CPS have not yet made the leap from one of a kind experiment to generally useful infrastructure, (3) lack of sufficient flexibility and modularity, and (4) the wealth of new and diverse users who would need better support tools, interfaces, and ease of use. Several problems/research directions in medical applications are also discussed in [11] and [12]. These problems include high assurance s/w, interoperability, context awareness, autonomy, security/privacy, and certifiability. CPS are also a promising approach to serve mission-critical applications (ensuring safety, security, and sustainability). Example mission-critical applications include emergency and management applications [13], [14]. In [4], Mueller discusses several security concerns in the power grid, and how immediate research is needed to protect the critical infrastructure to avoid cyber-physical attacks. Some researchers believe that we should put the security aspects in the forefront of our research objectives to avoid the major issues that researchers have faced after deployment/use of insecure Internet protocols ([15], [16], [17], and [18]). Finally, recent research has proposed protocols at road intersections to support autonomous vehicles [19].

VIII. CONCLUSION

In this work, we studied one important CPS application (traffic flow enhancement). We focused on what is needed to enable dynamic traffic light control for facilitating better traffic flow through intersections. Then, we studied the possible design approaches and proposed two protocols to achieve this goal. Our proposed techniques rely on input data from a sensor network on the road, and make control decisions either centrally at the traffic light, or in a distributed way using devices on the vehicles. Our analysis shows that our proposed techniques require low overhead in communication and processing. Simulations show that our proposed techniques provide significant benefits to the traffic flow metrics, especially in terms of vehicle throughput. Our future research plan is to study the use of CPS for traffic safety and pollution reduction. We also plan to study the effect of pedestrians and right/left turns on the performance of traffic flow at intersections.

REFERENCES

[1] E. A. Lee, "Cyber-Physical Systems - Are Computing Foundations Adequate?" NSF Workshop on Cyber-Physical Systems: Research Motivation, Techniques, and Roadmap, Austin, TX, Oct. 2006.

[2] M. Ferreira, R. Fernandes, H. Conceição, W. Viriyasitavat, O. K. Tonguz, "Self-organized traffic control," in Proceedings of the Seventh ACM International Workshop on VehiculAr InterNETworking (ACM VANET), Chicago, IL, June 2010, pp. 85-90.

[3] T. Gaugel, F. Shmidt-Eisenlohr, J. Mittag, H. Hartenstein, "A Change in Perspective: Information-Centric Modeling of Inter-Vehicle Communication," in Proceedings of the Eighth ACM International Workshop on VehiculAr InterNETworking (ACM VANET), Nevada, Sep. 2011.

[4] F. Mueller, "Challenges for Cyber-Physical Systems: Security, Timing Analysis and Soft Error Protection," in Proceedings of the National Workshop on High Confidence Software Platforms for Cyber-Physical Systems: Research Needs and Roadmap (HCSP-CPS), Nov 2006.

[5] T. Neudecker, N. An, O. K. Tonguz, T. Gaugel, J. Mittag, "Feasibility of virtual traffic lights in non-line-of-sight environments," in Proceedings of the Ninth ACM international Workshop on VehiculAr InterNETworking, Systems, and Applications (ACM VANET), United Kingdom, June 2012, pp. 103-106.

[6] V. Gradinescu, C. Gorgorin, R. Diaconescu, V. Cristea, and L. Iftode, "Adaptive Traffic Lights Using Car-to-Car Communication," in Proceedings of the IEEE Vehicular Technology Conference, Dublin, April 2007.

[7] P. Tabuada, "Cyber-Physical Systems: Position Paper," UCLA Technical Report (UCLA-TR-2012), 2012.

[8] T. Abdelzaher, "Toward an Architecture for Distributed Cyber-Physical Systems," Technical Report (UIUC-TR-2011), 2011.

[9] E. A. Lee, "CPS Foundations," in the ACM 47th Design Automation Conference (DAC), June 2010, pp. 737-742.

[10] R. Campbell, G. Garnett, and R. E. McGrath, "Cyber-Physical Systems: Position Paper," in NSF Workshop on Cyber-Physical Systems: Research Motivation, Techniques, and Roadmap, Austin, TX, Oct. 2006.

[11] L. Zhang, J. He, and W. Yu, "Challenges, Promising Solutions and Open Problems of Cyber-Physical Systems," in the International Journal of Hybrid Information Technology, Vol. 6, No. 2, March 2013.

[12] I. Lee, O. Sokolsky, S. Chen, J. Hatcliff, E. Jee, B. Kim, A. King, M. Mullen-Fortino, S. Park, A. Roederer, and K. Venkatasubramanian, "Challenges and Research Directions in Medical Cyber-Physical Systems," International Journal on Hybrid Information Technology, Vol. 6, No. 2, 2013.

[13] A. Baberjee, K. Venkatasubramanian, T. Mukherjee, S. Kumar, and S. Gupta, "Ensuring Safety, Security, and Sustainability of Mission-Critical Cyber-Physical Systems," Proceedings of the IEEE, Vol. 100, No. 1, January 2012.

[14] E. Gelenbe and F.-J. Wu, "Future Research on Cyber-Physical Emergency Management Systems," Future Internet, Vol. 5, 2013, pp. 336-354.

[15] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for Securing Cyber Physical Systems," in Proceedings of the Workshop on Future Directions in Challenges for Securing Cyber Physical Systems, July 2009.

[16] C. Neuman, "Challenges in Security for Cyber-Physical Systems," in the DHS Workshop on Future Directions in Cyber Physical Systems Security, July 2009.

[17] E. K. Wang, Y. Ye, X. Xu, S. M. Yiu, L. C. K. Hui, and K. P. Chow, "Security Issues and Challenges for Cyber Physical System," in the IEEE/ACM International Conference on Cyber, Physical, and Social Computing, 2010.

[18] H. Fawzi, P. Tabuada, S. Diggavi, "Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks," in the IEEE Transactions on Automatic Control, Vol. 59, No. 6, June 2014.

[19] R. Azimi, G. Bhatia, R. Rajkumar, and P. Mudalige, "STIP: Spatio-Temporal Intersection Protocols for Autonomous Vehicles," in IEEE/ACM International Conference on Cyber-Physical Systems, Germany, Apr. 2014, pp. 1-12.

Younis, Ossama; Moayeri, Nader.
"Cyber-Physical Systems: A Framework for Dynamic Traffic Light Control at Road Intersections."
Paper presented at IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar. April 3, 2016 - April 6, 2016.

Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid

Sérgio Câmara[1], Dhananjay Anand[2],
Victoria Pillitteri[2], and Luiz Carmo[1]

[1] National Institute of Metrology, Quality and Technology
Duque de Caxias, 25250-020, Rio de Janeiro, Brazil
{smcamara,lfrust}@inmetro.gov.br
[2] National Institute of Standards and Technology
Gaithersburg, MD 20899, USA
{dhananjay.anand,victoria.pillitteri}@nist.gov

Abstract. Multicast authentication of synchrophasor data is challenging due to the design requirements of Smart Grid monitoring systems such as low security overhead, tolerance of lossy networks, time-criticality and high data rates. In this work, we propose *inf*-TESLA, Infinite Timed Efficient Stream Loss-tolerant Authentication, a multicast delayed authentication protocol for communication links used to stream synchrophasor data for wide area control of electric power networks. Our approach is based on the authentication protocol TESLA but is augmented to accommodate high frequency transmissions of unbounded length. *inf*-TESLA protocol utilizes the Dual Offset Key Chains mechanism to reduce authentication delay and computational cost associated with key chain commitment. We provide a description of the mechanism using two different modes for disclosing keys and demonstrate its security against a man-in-the-middle attack attempt. We compare our approach against the TESLA protocol in a 2-day simulation scenario, showing a reduction of 15.82% and 47.29% in computational cost, sender and receiver respectively, and a cumulative reduction in the communication overhead.

Keywords: Multicast authentication, Smart Grid, synchrophasors, Wide Area Monitoring Protection and Control

1 Introduction

Smart Grids are large critical cyber-physical infrastructures and are being transformed today with the design and development of advanced real-time control applications [11]. The installation of Phasor Measurement Units (PMUs) as part of world-wide grid modernization is an example of major infrastructure investments that require secure standards and protocols for interoperability [1].

PMUs take time-synchronized measurements of critical grid condition data such as voltage, current, and frequency at specific locations that are used to provide wide area visibility across the grid.The synchrophasor data aggregated

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

from multiple PMUs are used to support real-time analysis, planning, corrective actions, and automated control for grid security and resiliency. Currently, high-speed networks of PMUs are being used for Wide Area Monitoring Protection and Control (WAMPAC) applications to provide situational awareness in the Eastern and Western Interconnection of North America, in China, Canada, Brazil and across Europe [11]. Before the installation of PMUs, the lack of wide-area visibility is one of the factors that prevented early fault identification of the 2003 Northeast America and 2003 Italy blackouts [21] [9]. Malicious PMU data or deliberate attacks could result in inaccurate decisions detrimental to grid safety, reliability, and security, that said, PMUs need information authentication and integrity, while confidentiality may be considered optional.

Authentication schemes in the Smart Grid must be able to efficiently support multicast. Current standard solution, suggested by IEC 62351 [5], comprises HMAC authentication algorithm for signing the synchrophasors. However, sharing only one symmetric key across a multicast group cannot guarantee adequate security, and this approach suffers from the scalability problem. The use of asymmetric cryptography and digital signatures for multicast authentication raises concerns about the impact on cost and microprocessor performance. One-Time Signature schemes can enable multicast authentication, however they suffer from communication and storage overhead, and complicated key management [24].

Although some previous literature works assume, in general, that delayed authentication is not suitable for real-time applications [7] [8], such method is still eligible for some monitoring and control applications that permit relatively larger delay margins (e.g. wide-area oscillation damping control application) [25]. For more considerations on this topic, see Section 2. Moreover, delayed authentication presents advantages over cited issues by supporting multicast data streaming, symmetric and lightweight cryptography, corrupt data and attack detection. Also it allows scalable solutions and key management, tolerates packet loss, and provides low communication overhead and high computational efficiency.

The primary objective of this work is to propose a multicast delayed authentication protocol called *inf*-TESLA in order to provide measurement authentication in a WAMPAC application within the Smart Grid. Also, we design the Dual Offset Key Chains mechanism which is used by our protocol to generate the authenticating keys and to provide long-term communication without the need of key resynchronization between the sender and receivers. A description of two different modes for disclosing keys and a demonstration of a man-in-the-middle attack attempt against out mechanism are also provided.

Section 2 presents an overview of the network architecture used for wide area aggregation of PMU data as well as some delay constraints and authentication infrastructure. In Section 3 we discuss prior work in the area of packet based authentication protocols for streaming communication, and then in Section 4 we present the *inf*-TESLA protocol and describe the Dual Offset Key Chains mechanism along with its security properties and conditions. In Section 5 we evaluate our approach against the original TESLA protocol. Finally, we summarize our results and propose future works in Section 6.

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

2 Scenario Characteristics

The network architecture considered for this work is as follows. Each communication link in the infrastructure comprises one PMU sender node S capable of multicasting packets to m receivers R_k applications, where $1 \leq k \leq m$. PMU S sends time-stamped synchrophasor data packets at a rate of 10 to 120 packets per second and that can be dropped in the way to the receivers. The network has several n intermediate nodes between S and $R_k, n > 0$, called Phasor Data Concentrators (PDCs). PDCs can chronologically sort received synchrophasors as well as aggregate, repackage and route data packets to the set of higher level PDCs (Super PDCs). When packets are missing or lost, PDCs may (with due indication) interpolate measurements in order to retain the communication link.

There are different wide-area monitoring and control applications that consume synchrophasor data and have different time delays and quality requirements. For instance, Situational Awareness Dashboard, Small-Signal Stability Monitoring, and Voltage Stability Monitoring/Assessment accept up to 500 milliseconds in communication latency, other applications such as Long-term stability control, State Estimation, and Disturbance Analysis Compliance can handle up to 1000 ms. For the entire list, see [20].

Zhu *et al.* [25] simulates the latency for monitoring applications over the Smart Grid network architecture and obtained results within a range of 150–220 ms. For centralized control applications, the latency was well below 500 ms. From the delayed authentication perspective, the minimum delay of the authentication confirmation by R_k is approximately twice the latency of the network. Still, delayed authentication protocols are able to attend the requirements for the above cited applications.

When utilizing multicast communication, IEC 61850-90-5, the standard for communication networks and systems for power utility automation, requires a Key Distribution Center (KDC), which provides the symmetric key coordination between S and R_k. We assume that each S is its own KDC, which is also endorsed by the standard. Furthermore, as our scheme demands that S prove its identity to R_k once during communication initialization, each receiver is required to validate a digital signature from S and maintaining a copy of its public key certificate. For this purpose, we assume that a Public-Key Infrastructure (PKI) is also available.

2.1 Security Considerations

We assume that attacks are accordingly aligned, via a man-in-the-middle, to either manipulate data values or masquerade as a legitimate PMU. Using the attack model from [23], the adversary is not limited by network bandwidth and has full control to drop, resend, capture and manipulate packets. Although his computational resources can be large, it is not unbounded and he cannot invert a pseudorandom function with non-negligible probability. Each receiver R_k is able to authenticate both the content and source of synchrophasor payloads after a delay of d_{NMax} using our delayed authentication scheme presented in Section 4.

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

However, if a packet fails authentication at time t, then an attack that has been active and undetected since $t - d_{NMax}$ represents the maximum threat exposure.

The security primitives used throughout this paper are as follows:

- *One-way hash function* H operates on an arbitrary length input message M, returning $h = H(M)$. H can be implemented with SHA-2 family algorithms.
- *Message Authentication Code* $MAC(K, M)$ provides a tag that can verify authenticity and integrity of message M given a shared key K. $HMAC(K, M)$ is a specific construction which includes an underlying cryptographic hash function to create the authenticating tag.
- *Hash chain* $H^n(M)$ denotes n successive applications of cryptographic hash function H to message M.

3 Related Work

Multicast authentication is an active research field in recent years and has been applied to a wide range of applications. In Smart Grids, it is being used for monitoring, protection and information dissemination [24]. In this section, we review all the TESLA-based multicast authentication schemes and other multicast authentication schemes used for electrical power systems.

To address the challenge of continuous stream authentication for multiple receivers on a lossy network, Timed Efficient Stream Loss-tolerant Authentication (TESLA) was introduced by Perrig et al [14]. Based on the Guy Fawkes protocol [2] and requiring loose time synchronization between the senders and receivers, TESLA is a broadcast authentication protocol considering delayed disclosure of keys used for authentication of previous sent messages and packet buffering by the receiver. This protocol supports fixed/dynamic packet rate and delivers packet loss robustness and scalability. Benefits of TESLA include a low computation overhead, low per-packet communication overhead, arbitrary packet loss is tolerated, unidirectional data flow, high degree of authenticity and freshness of data. Further work proposed several modifications and improvements to TESLA, allowing receivers to authenticate packets upon arrival, improved scheme scalability, reduction in overhead, and increased robustness to denial-of-service attacks [13].

Studer et al. describe TESLA++ [19], a modified version of TESLA resilient to memory-based DoS attacks. They combine TESLA++ and ECDSA signatures to build an authentication framework for vehicular ad hoc networks.

μTESLA [17] adapts TESLA to make it practical for broadcast authentication in severely resource-constrained environments; like sensor networks. Some of these adaptations include the use of only symmetric cryptography mechanisms, less frequent disclosure of keys and restriction on the number of authenticated senders. Liu and Ning [10] reduce the overhead needed for broadcasting key chain commitments and deal with DoS attacks. Their Multilevel μTESLA protocol considers different levels of key chains to cover the entire lifespan of a sensor.

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

Other methods include the One-Time Signatures family which gained popularity recently and is applicable to multicast authentication and also for WAMPAC applications. The author in [12] describes a one-time signature based broadcast authentication protocol based on BiBa. BiBa uses one-way functions without trapdoors and exploits the birthday paradox to achieve security and verification efficiency. Its drawbacks include a large public key and high overhead for signature generation.

HORS [18] is described by Reyzin et al. as an OTS scheme with fast signing and signature verification using a cryptographic hash function to obtain random subsets for the signed message and for verifying it, but it still suffers from frequent public key distribution. TSV [8] multicast authentication protocol generates smaller signatures than HORS and has lower storage requirement at the cost of increased computations in signature generation and verification. TSV+ [7], a patched version of TSV, uses uniform chain traversal and supports multiple signatures within an epoch. SCU [22] is a multicast authentication scheme designed for wireless sensor networks and SCU+ [7] adapts it for power systems using uniform chain traversal as well. TV-HORS [23] uses hash chains to link multiple key pairs together to simultaneously authenticate multiple packets and improves the efficiency of OTS by signing the first l bits of the hash of the message. As a downside, TV-HORS has a large public key of up to 10 Kbytes.

4 Proposed Solution

In this section, we propose *inf*-TESLA, a new TESLA based scheme that improve its overall performance. At first, we review TESLA to give some background and then present our scheme.

4.1 TESLA

Timed Efficient Stream Loss-tolerant Authentication (TESLA) [14] [13] [15] [16] is a broadcast authentication protocol with low communication and computation overhead, tolerates packet loss and needs loose time synchronization between the sender and the receivers.

TESLA relies on the delayed disclosure of symmetric keys, therefore the receiver must buffer the received messages before being able to authenticate them. The keys are generated as an one-way chain and are used and disclosed in the reverse order of their generation. At setup time, the sender must first set n as the index of the first element K_n. For generating the key chain, the sender picks a random number for K_n and using a pseudo-random function f, he constructs the one-way function $F : F(k) = f_k(0)$. So, the sender generates recursively all the subsequent keys on the chain using $K_i = F(K_{i+1})$. By that, the last element of the chain is $K_0 = F^n(K_n)$, and all other elements could be calculated using $K_i = F^{n-i}(K_n)$.

Each K_i looks pseudo-random and an adversary is unable to invert F and compute any K_j for $j > i$. In the case of a lost packet containing K_i, a receiver

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

can calculate K_i given any subsequent packet containing K_j, where $j < i$, since $K_j = F^{i-j}(K_i)$. As a result, TESLA tolerates sporadic packet losses.

The stream authentication scheme of TESLA is secure as long as the security condition holds: A data packet P_i arrived *safely*, if the receiver can unambiguously decide, based on its synchronized time and maximum time discrepancy, that the sender did not yet send out the corresponding key disclosure packet P_j.

TESLA also supports both communication with fixed or dynamic packet rate. For fixed rate, the sender discloses the key K_i of the data packet P_i in a later packet P_{i+d}, where d is a delay parameter set and announced by the sender during setup phase. The sender determines the delay d according to the packet rate r, the maximum tolerable synchronization uncertainty δ_{tMax} and the maximum tolerable network delay d_{NMax}, setting $d = \lceil(\delta_{tMax} + d_{NMax})r\rceil$. In this mode, the scheme can achieve faster transfer rates. For dynamic rate, the sender pick one key per time interval T_{int}. Each key is assigned to a uniform interval of duration T_{int}, T_0, T_1 ,..., T_n, that is, key K_i will be active during the time period T_i. The sender uses the same key K_i to compute the MAC for all packets which are sent during T_i, on the other hand, all packets during T_i disclose the key $K_{i-d'}$. In this case, $d' = \lceil(\delta_{tMax} + d_{NMax})/T_{int}\rceil$. We use the designation d and d' for fixed and dynamic rates respectively.

For each new receiver that joins the communication network, the sender initially creates an authenticated synchronization packet. This packet contains parameters such as interval information, the disclosure lag and also a disclosed key value - which is a commitment to the key chain. The sender digitally signs this packet to each new receiver before starting the streaming communication.

4.2 *inf*-TESLA

inf-TESLA, short for infinite TESLA, is a multicast authentication protocol based on TESLA suitable for use in long term communication at high packet rates. As in TESLA, *inf*-TESLA relies on the strength of symmetric cryptography and hash functions and on the delayed disclosure of keys as a means to authenticate messages from the sender. Also, it requires only loose time synchronization between the sender and the receiver and can operate under both dynamic and fixed packet rates.

By using fixed packet rate mode, there is no need for setting specific time intervals for MACing and disclosing keys. Each autheticating key is used once for the actual message and disclosed d packets later. Although this operational mode can achieve maximum speed on authenticating previous packets, it has a drawback of quickly consuming the authenticating key chain, depending on the frequency of the packets.

Since we use one-way hash functions to build independent key chains, every time one of the key chains comes to an end (meaning that it was fully used in the authentication process) the sender must automatically build, store and utilize a new key chain in its place. In the original TESLA protocol, a sender would have to reassign a new synchronization packet as the current key chain comes to an

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

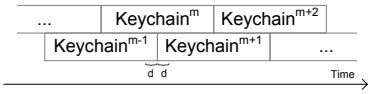


Fig. 1. An illustration of dual offset key chains as used for *inf*-TESLA.

end, inflicting non-negligible network and computational overhead by digitally signing a synchronization packet at the end of each key chain.

inf-TESLA addresses this issue by using the Dual Offset Key Chains mechanism. This mechanism uses a pair of keys for each message and guarantees continuity of the multicasting authentication process without the need for signing and sending a new synchronization packet. The mechanism creates two offset key chains so that a pair of active key chains are always available and, as the main principle, a key chain m always straddles the substitution of key chain $m-1$ with $m+1$. Figure 1 illustrates the Dual Offset Key Chains mechanism by which key chain m supports the substitution of key chain $m-1$ for key chain $m+1$ without the need for resynchronization. A detailed description of the Dual Offset Key Chains mechanism is presented on Section 4.2.

The overall initialization setup is similar to TESLA. Before the data streaming begins, the sender first determines some fundamental information about the network status, d_{NMax}), and time synchronization, δ_{tMax}, and builds its first two key chains. We assume that both sender and receiver are time synchronized by a reliable time protocol (e.g. PTP). After that, the sender S chooses the delay parameter d (Section 4.1) that will base the decision of the receiver R_k to either accept a packet from S. This condition is **Security Condition-1** for *inf*-TESLA.

For bootstrapping each new receiver, S constructs and sends the synchronization (commitment) packet to the new incomer. For a dynamic packet rate, this packet contains the following data [13]: the beginning time of a specific interval T_j along with its id I_j, the interval duration T_{int}, the key disclosure delay d', a commitment to the key chain K_i^m and key chain K_i^{m+1} ($i < j - d'$ where j is the current interval index).

For a fixed packet rate r, let j_1 and j_2 be the current key from key chains m and $m+1$ respectively. The synchronization packet contains: delay d and the commitment for the key chains $K_{i_1}^m$ and $K_{i_2}^{m+1}$ ($i_1 < j_1 - d$ and $i_2 < j_2 - d$). We will focus on fixed packet rate in this paper for the sake of brevity and convenience of notation. While a fixed packet rate is potentially more likely for the streaming applications we address, our approach is compatible with both dynamic and fixed rates.

Dual Offset Key Chains mechanism. The Dual Offset Key Chains mechanism enables continuity in streaming authentication without the periodic resynchronization between S and $R_k \in R$ required by TESLA. Two key chains, offset in alignment, are used simultaneously by the mechanism to authenticate mes-

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

sages. For every packet, there are always two active key chains and, from each chain, one non-used key available for MACing.

For constructing the two key chains, first the sender chooses n, the total number of elements on a single key chain. Let l^m be the current number of remaining elements on the key chain m. Here we assume that all created keys are deleted just after being used for authenticating messages. Let M be the maximum available memory for storing the key chains, assuming that M is big enough for storing two key chains, m and $m+1$, at any time. The value of n must be chosen accordingly to the following constraints: (i) $n \geq l^{m-1} + 2(d+1)$ and (ii) $n \leq \frac{M}{2} + d$.

The the first constraint sets a minimum value for n, that is the minimum initial size of a key chain. During the initialization setup of the first receiver synchronization, we consider $l^{m-1} = 0$ for constructing the first key chain. The second constraint restricts the maximum number of elements in a key chain. If a key chain m does not meet this limit, key chain $m+1$ will not be long enough to meet the security condition for the key chain exchange procedure (see Section 4.2). In practice, it may not be feasible to calculate a whole key chain in the time taken to send two data packets and so S may compute and store key chain $m+1$ well before the end of key chain $m-1$.

A packet P_j sent by S is formed by the following data $P_j = \{M_j, i_1, i_2, K^m_{i_1-d}, K^{m+1}_{i_2-d}, MAC(K^m_{i_1}||K^{m+1}_{i_2}, M_j)\}$. Every packet carries the actual message M_j, the current sequence number of each key chain i_1 and i_2, the disclosed authenticating keys $K^m_{i_1-d}$ and $K^{m+1}_{i_2-d}$ (discussed later in Section 4.2) and the MAC of the message resultant from an operation that uses the concatenation of current keys from both key chains. In particular, at the beginning of a key chain K^m, the notation K^m_{i-d} may refer to the last keys in the key chain K^{m-1}.

Disclosure of keys. *inf*-TESLA has two modes of operation for disclosing keys: **2-keys** and **Alternating**. In the 2-keys mode (or standard mode, as previously described), each packet P_j discloses two authentication keys, one from each key chain, for the same message M_i, that is packet P_j has the following information, $P_j \rightarrow K^m_{i_1}, K^{m+1}_{i_2}$.

The Alternating mode discloses one key from each key chain alternatively in each data packet. Formally, two consecutive packets would have the following information about keys, $P_j \rightarrow K^m_{i_1}$ and $P_{j+1} \rightarrow K^{m+1}_{i_2+1}$, where indexes i_1 and i_2 correspond to the keys of both key chains to be disclosed in the same data packet in 2-keys mode of operation. Figure 2 shows the key chains in time and the two modes for disclosing keys. In Section 5, we present a more detailed comparison of these two modes in relation to communication overhead, computational cost and authentication delay.

The disclosure delay d for the keys is directly affected by the maximum tolerable network delay d_{NMax}, so each receiver R_k will present a different delay value. Sender S must set d as the largest expected delay in order to meet security condition-1.

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

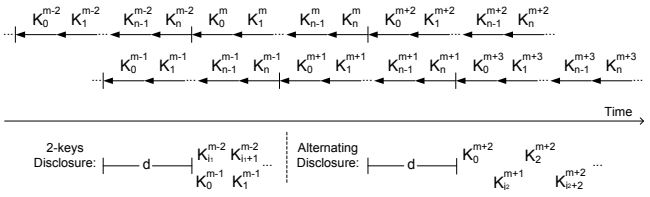


Fig. 2. Two modes for disclosing keys: 2-keys and alternating.

Dual Offset Key Chains mechanism security. Key chain security is based on the widely used cryptographic primitive: the one-way chain. One-way chains were first used by Lamport for one-time password [6] and has served many other applications in the literature.

The **Security Condition-2** for *inf*-TESLA concerns the key chain exchange procedure. This condition states that both key chains cannot be substituted within a time interval d/r (or within d packets). If this happens, the receiver must drop the following packets and request for resynchronization with the sender. This protocol restriction assures the authentication inviolability of *inf*-TESLA and must be observed at all times by the receiver. The receiver is solely responsible for monitoring the key chain exchange procedure and accepting, or rejecting, the new key chain.

In Figure 3, we show an example of a man-in-the-middle attack attempt on the Dual Offset Key Chains mechanism and the importance of the security condition-2. For this example, we consider $d = 9$ as minimum number packets the sender has to wait to disclose a key, the last element $n = 50$ for all key chains, and the asterisk symbol indicates an item maliciously inserted by the attacker. The packets are presented without indices "i" for cleaner presentation.

We first illustrate how this attack can work on a single key chain mechanism without commitment packets as follows: When the attacker senses a change in the key chain by testing every disclosured key *(a)*, he inserts M_0^* as the first manipulated message and MACs it using the first element K_0^* of a forged key chain of his own. The attacker continues faking the messages and its MACs till the last authentic key used for MACing is disclosured. After that point, the attacker is able to take complete control of the communication without being detected (b). For the second part of Figure 3, the same attack is attempted against our mechanism. Also the attacker is able to sense when a disclosured key chain comes to an end and can also substitute the messages and the MACs in the packets. However, when he tries to take complete control of the key chain by forcing the forged key K_0^{**} over the key chain $m = 2$, this indicates for the receiver a violation of the security condition-2 for the key chain exchange procedure.

Another concern is how many consecutive packets could be lost by the receiver without actually being an attack. Following the security condition for key chain substitution, there must not be two different key chain substitutions

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

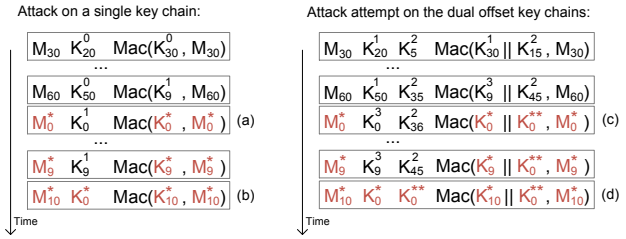


Fig. 3. Example of a man-in-the-middle attack on a single hash chain without commitment and on the Dual Offset Hash Chains mechanism.

within a period of d/r, so the receiver must be aware that the limit for consecutive packets lost is at maximum d. If, for some reason, more than d packets are lost/dropped, the receiver must assure that the following disclosed keys are authentic elements of at least one of the existing key chains, otherwise the receiver will not be able to authenticate any of the next received packets. From this point, the receiver must refuse this stream and request for a new synchronization with the sender.

Another security issue can occur when the last key K_n in the key chain's sequence is lost, that can cause a total lack of authentication of a previous packet P_n. When some K_i is lost, it can be computed from any subsequent key in the key chain through function F (Section 4.1), however when $i = n$ there is no subsequent key. This issue can be extended for the last d elements of the key chain, meaning that in this scenario some packets may not be authenticated and then must be dropped by the receiver. For the Alternating mode for disclosing keys, the receiver would drop d+1 packets in the worst case. This issue concerning the last keys of the key chain is a vulnerability of the original TESLA as well.

Elaborate attacks, like selective drop of packets, can cause even more authentication delay without being noticed. For instance, in the case of the Alternating keys disclosure mode, one attacker can induce an alternating drop of packets preventing the sender to authenticate some sequential packets. To mitigate these attacks, the receiver must set an upper limit for the maximum number of non authenticated packets to ignore before resynchronizing with the sender.

5 Evaluation against TESLA

For the following comparison evaluation, we check for communication overhead, authentication delay and computational cost on a long term communication for each of the following schemes: original TESLA, *inf*-TESLA 2-keys (two disclosured keys per packet) and *inf*-TESLA alt (alternating key chain disclosure). Due to PMUs' operational settings, we are only considering a fixed packet rate mode. Also, we assume the following constraints for the simulation:

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

Table 1. Communication overhead.

	Formula
TESLA (fixed)	$C * (sKey + sSig) + P * (sKey + sMac)$
Inf-TESLA 2-keys	$2 * sKey + sSig + P * (2 * sKey + sMac)$
Inf-TESLA alt	$2 * sKey + sSig + P * (sKey + sMac)$
	2-day simulation (MBytes)
TESLA (fixed)	331,825
Inf-TESLA 2-keys	497,664
Inf-TESLA alt	331,776

- Phasor data frame size of 60 bytes, according to the C37.118 standard [25], over UDP transport layer protocol.
- Pseudo-Random function and HMAC function implementation as HMAC-SHA-256-128. Both HMAC key size and HMAC tag size (truncated) of 128 bits.
- Digital signature implemented as ECDSA over GF(p) of 256 bits. Although TESLA considers RSA signatures, for comparison purposes we use ECDSA signatures. The keys and signatures sizes are based on the NIST SP 800-131A [3] for recommendations on use of cryptographic algorithms and key lengths.
- Maximum number of keys n that can be stored at a time in the cache memory of a device is 10,000 keys.
- Sender's packet rate (frequency) of 60 packets/sec.
- Simulation testing time of 2 days. Past references [7] established a baseline of 1024 key chains for evaluating the one-time signature multicast schemes. However, as *inf*-TESLA must build approximately 4 times the number of key chains as TESLA for the same number of packets, comparisons are done for fixed simulation duration rather than number of key chains.

Table 1 shows the formulas to calculate all security related communication overhead of each of the 3 schemes. Let C be the number of commitments (signed packets), P the total number of transmitted packets and $sKey$, $sMac$ and $sSig$ be the size of a cryptographic key, the size of the MAC tag and the size of a signature tag respectively. *inf*-TESLA 2-keys presents the higher communication overhead due to two disclosed keys per packet, while TESLA and *inf*-TESLA alt present a slightly, but negligible, difference on the overhead during two days of communication.

For calculating the computational cost overhead of each scheme, we use the formulas shown in Table 2. The processing cost in cycles per each operation of hashing, macing, signing and verifying is represented by $cHash$, $cMac$, $cSig$ and $cVer$ respectively. From the graph in Figure 4, we can observe the higher computational cost of the sender and receiver operating TESLA over *inf*-TESLA, due to constant signing and verification operations.

For two days of simulation in this configuration, a sender running TESLA protocol on fixed packet rate mode has to sign up to 1036 commitment packets and spends on average 0.373117 gigacycles/hour, while running *inf*-TESLA he would spend 0.314087 gigacycles/hour of operation, which means a reduction

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

Table 2. Computational cost calculation.

	Sender
TESLA (fixed)	$C * cSig + P * (cMac + cHash)$
Inf-TESLA (both)	$cSig + 2 * P * (cMac + cHash)$
	Receiver
TESLA (fixed)	$C * cVer + P * (cMac + cHash)$
Inf-TESLA (both)	$cVer + 2 * P * (cMac + cHash)$

of 15.82% in computational cost for the sender. On the receiver side, a TESLA receiver spends in average 0.596289 gigacycles/hour, while *inf*-TESLA needs 0.314303 gigacycles/hour, meaning a reduction of 47.29% in computational cost for the receiver. All values of cycles/operation of the security primitives are referenced from the Crypto++ Library 5.6.0 Benchmarks [4].

Although the alternating keys disclosure mode showed good results on the two previous evaluations, this mode increases the authentication delay of a packet P_i by one packet. That is because the second key needed for authenticating P_i, i.e. $K_{i_2}^{m+1}$, will only be disclosed on P_{j+1} where $j > i + d$. Also, if P_{j+1} happens to be lost, the authentication of P_i will be only achieved when receiver has the disclosed key included in P_{j+3}. On both other schemes, the authentication of a packet P_i is normally achieved after receiving P_j, $j > i + d$, and if P_j is lost, the missing keys can be recovered from the contents in P_{j+1}. Also regarding authentication delay evaluation, necessary time overhead for generation and verification of digital signatures during key chains exchange may affect TESLA's continuous flow on higher frequencies of streamed data.

Although TESLA protocol is an efficient protocol and has low security overhead, it was not originally designed for long-term communication. We observe that *inf*-TESLA, in alternating disclosure mode, can deliver a slightly lower

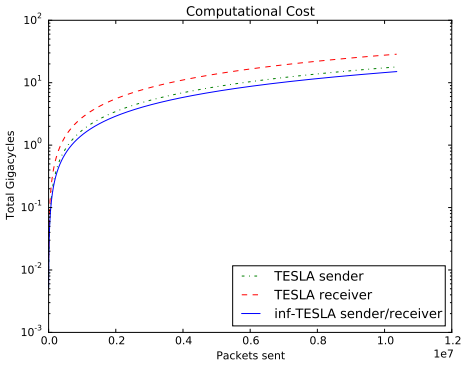


Fig. 4. Computational cost for TESLA and *inf*-TESLA over 2 days of streaming data.

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

communication overhead and, for both modes, result in a significant reduction in computational overhead over the original protocol. In general, *inf*-TESLA scheme also provides great suitability for key storage and computational constrained devices, such as in Wireless Sensor Networks (WSNs).

6 Conclusion

In this work, we present *inf*-TESLA, a multicast delayed authentication protocol for streaming synchrophasor data in the Smart Grid, suitable for long-term communication and high data rates scenarios. To authenticate messages from the sender, *inf*-TESLA uses two keys to generate the MAC of the message and discloses both keys after a time frame d/r, on a fixed packet rate of operation.

We also design the Dual Offset Key Chains mechanism to produce the authenticating keys and provide a long-term communication without the need of frequently signing resynchronization packets containing commitments to the new key chains, which ensures continuity of the streaming authentication. We prove our mechanism against a man-in-the-middle attack example and describe the security conditions that must be observed at all times by the receiver. *inf*-TESLA enables two different modes for disclosing keys, 2-keys (or standard) and Alternating keys. We present a comparison between this two modes against TESLA within a WAMPAC application, and our protocol shows even more efficiency when compared to the original. Although the Alternating key disclosure mode increases the authentication delay by one packet, it provides less impact on communication overhead and a reduction of 15.82% and 47.29%, sender and receiver respectively, in computational cost during operational time. Generally, *inf*-TESLA shows promise and suitability for key storage and computational constrained devices.

In future work, we intend to do a further analysis on the trade-off between key storage size in devices and protocol performance, and on the possible (minimum/maximum/average) values for the authentication delay by simulating our protocol in a WAMPAC network.

References

1. Greer et al., C.: NIST Framework and Roadmap for Smart Grid Interoperability Standards. Tech. rep., NIST (2014)
2. Anderson, R., Bergadano, F., Crispo, B., Lee, J.H., Manifavas, C., Needham, R.: A new family of authentication protocols. ACM SIGOPS Operating Systems Review 32, 9–20 (1998)
3. Barker, E., Roginsky, A.: Recommendation for transitioning the use of cryptographic algorithms and key lengths. SP 800-131A Transitions (2011)
4. Dai, W.: Crypto++ 5.6. 0 benchmarks. Website at http://www.cryptopp.com/benchmarks.html (2009)
5. International Electrotechnical Commission: IEC TS 62351-1 Power systems management and associated information exchange - Data and communications - Part 1: Communication network and system security-Introduction to security issues (2007)

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

6. Lamport, L.: Password authentication with insecure communication. Communications of the ACM 24(11), 770–772 (1981)
7. Law, Y.W., Gong, Z., Luo, T., Marusic, S., Palaniswami, M.: Comparative study of multicast authentication schemes with application to wide-area measurement system. Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security p. 287 (2013)
8. Li, Q., Cao, G.: Multicast authentication in the smart grid with one-time signature. IEEE Transactions on Smart Grid 2, 686–696 (2011)
9. Liscouski, B., Elliot, W.: Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. A report to US Department of Energy 40(4) (2004)
10. Liu, D., Ning, P.: Multilevel μTESLA: Broadcast Authentication for Distributed Sensor Networks. ACM Trans. Embed. Comput. Syst. 3, 800–836 (2004)
11. Patel, M., Aivaliotis, S., Ellen, E.: Real-time application of synchrophasors for improving reliability. NERC Report, Oct (2010)
12. Perrig, A.: The BiBa one-time signature and broadcast authentication protocol. Proceedings of the 8th ACM conference on Computer and Communications Security p. 28 (2001)
13. Perrig, A., Canetti, R., Song, D.: Efficient and secure source authentication for multicast. Proceedings of the Internet Society Network and Distributed System Security Symposium pp. 35–46 (2001)
14. Perrig, A., Canetti, R., Tygar, J.D., Song, D.: Efficient Authentication and Signing of Multicast Streams over Lossy Channels. Proceedings of the IEEE Symposium on Security and Privacy 28913, 56–73 (2000)
15. Perrig, A., Canetti, R., Tygar, J., Song, D.: The TESLA broadcast authentication protocol. CryptoBytes Summer/Fall, 2–13 (2002)
16. Perrig, A., Song, D., Canetti, R., Tygar, J., Briscoe, B.: Timed efficient stream loss-tolerant authentication (TESLA): Multicast source authentication transform introduction. The Internet Society RFC:4082, 1–22 (2005)
17. Perrig, A., Szewczyk, R., Tygar, J., Wen, V., Culler, D.E.: Spins: Security protocols for sensor networks. Wireless networks 8(5), 521–534 (2002)
18. Reyzin, L., Reyzin, N.: Better than BiBa: Short One-Time Signatures with Fast Signing and Verifying. Information Security and Privacy 2384, 1–47 (2002)
19. Studer, A., Bai, F., Bellur, B., Perrig, A.: Flexible, extensible, and efficient VANET authentication. Journal of Communications and Networks 11, 574–588 (2009)
20. Tuffner, F.: Phasor Measurement Unit Application Data Requirements. Tech. rep., Pacific Northwest National Laboratory (2014)
21. UCTE: Final Report of the Investigation Committee on the 28 September 2003 Blackout in Italy. Tech. Rep. April, Union for the Coordination of the Transmission of Electricity (2004)
22. Ugus, O., Westhoff, D., Bohli, J.M.: A rom-friendly secure code update mechanism for wsns using a stateful-verifier τ-time signature scheme. In: Proceedings of the second ACM conference on Wireless network security. pp. 29–40. ACM (2009)
23. Wang, Q., Khurana, H., Huang, Y., Nahrstedt, K.: Time valid one-time signature for time-critical multicast data authentication. Proceedings - IEEE INFOCOM pp. 1233–1241 (2009)
24. Wang, W., Lu, Z.: Cyber security in the Smart Grid: Survey and challenges. Computer Networks 57(5), 1344–1371 (April 2013)
25. Zhu, K., Nordstrom, L., Al-Hammouri, A.: Examination of data delay and packet loss for wide-area monitoring and control systems. In: Energy Conference and Exhibition (ENERGYCON), 2012 IEEE International. pp. 927–934 (Sept 2012)

Camara, Sergio; Anand, Dhananjay; Pillitteri, Victoria; Carmo, Luiz.
"Multicast Delayed Authentication For Streaming Synchrophasor Data in the Smart Grid."
Paper presented at 31st International Conference on ICT Systems Security and Privacy Protection - IFIP SEC 2016, Ghent, Belgium. May 30, 2016 - June 1, 2016.

2016 IEEE 27th International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC): Services Applications and Business

PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps

Nader Moayeri*, Mustafa Onur Ergin†, Filip Lemic†, Vlado Handziski†, and Adam Wolisz†
*National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA
†Telecommunication Networks Group (TKN), Technische Universität Berlin, Berlin, Germany
Email: moayeri@nist.gov, {ergin,lemic,handziski,wolisz}@tkn.tu-berlin.de

***Abstract*—This paper describes a major effort to collect smartphone data useful for indoor localization. Data has been collected with four Android phones according to numerous scenarios in each of four large buildings. The data includes time-stamped traces of various environmental, position, and motion sensors available on a smartphone, Wi-Fi and cellular signal strengths, and GPS fixes, whenever available. Quantitative evidence is presented to validate the collected data and attest to its quality. This unique, extensive data repository is made available to the R&D community through the PerfLoc web portal to facilitate development of smartphone indoor localization apps and to enable performance evaluation of such apps according to the ISO/IEC 18305 international standard in the near future.**

I. Introduction

The Global Positioning System (GPS) has been a phenomenal success with applications in a wide range of domains, but it does not work inside buildings / structures and in urban canyons. The next frontier is to provide localization and tracking capability indoor as a key technology enabler for Location Based Services (LBS), which is anticipated to be a multi-billion dollar market. Indoor localization and tracking has applications in many areas, including emergency response for better coordination of operations and to save lives of first responders and civilians; E-911; military operations, such as search and clear operations and prevention of friendly fire; tracking in underground coal mines, particularly in the aftermath of explosions and roof collapses; asset and personnel tracking in warehouse, hospitals, and factories; tracking of children on school grounds and the elderly; offender tracking; navigation in museums, shopping malls, and large office buildings; urban search and rescue in the aftermath of natural / man-made disasters; and many applications related to the Internet of Things (IoT).

Yet, lack of standardized testing has been an impediment to the wide adoption and deployment of indoor Localization and Tracking Systems (LTSs). It is not possible to compare the performance of various systems presented in academic conferences or developed by industry, because they are evaluated in different environments and according to different and typically inadequate testing criteria and methodologies. This has made it difficult to set minimum performance requirements[1] for indoor LTSs. Consequently, the user community has often been unable to ascertain whether a given indoor LTS meets its requirements. There has been a strong demand by the user community for development of standardized testing procedures for indoor LTSs, which would also help the industry to improve the performance and effectiveness of their indoor LTS products.

In response to this demand, NIST led the development of the international standard ISO/IEC 18305, Test and evaluation of localization and tracking systems [1]. The primary focus of ISO/IEC 18305 is on indoor localization, which is a much harder problem than its outdoor counterpart, for which GPS and to a lesser extent the use of terrestrial cellular technology are the dominant solutions. The standard deals with a challenging test and measurement problem, because the performance of such systems is affected by a wide range of factors, including construction material, size and floor plans of buildings; mobility mode of the entity to be localized and tracked (stationary, walking, running, sidestepping, walking backwards, crawling on the floor, transportation on a cart or forklift, transportation in elevators, etc.); availability of coordinates of the boundary (footprint) of buildings; availability of floor plans and heights of different floors of buildings. While it might be possible to test the "components" of an LTS in a laboratory setting, the proper way to test the "system" is to do so in several large buildings, including high rises and subterranean structures, according to a variety of mobility scenarios. In addition, it is important to determine how many entities the system can localize and track. LTSs can vary significantly in terms of the assumptions under which they can operate, cost, size, weight, battery life, electromagnetic compatibility, intrinsic safety, etc. Therefore, one has to be careful when comparing the performance of different LTSs to ensure fairness. Tailoring the testing procedure to the LTS is not scalable, because there are many LTSs available. ISO/IEC 18305 treats the LTS under test as a black box and yet it provides a comprehensive testing methodology that adequately tests LTSs.

This paper focuses on smartphone indoor localization apps, which can use only the sensors available on a smartphone and the Radio Frequency (RF) signals that a smartphone can receive. In contrast, a general LTS can use other sensors and RF technologies, such as Ultra Wideband (UWB) ranging, angle of arrival estimation, and LiDAR. However, with billions of smartphones in use around the world, the smartphone platform is very important. The main objective of this effort is to create a level playing field for comparison of indoor localization apps. We are doing this by (i) making available to

The work of the first author has been supported by the Public Safety Communications Research (PSCR) Division at NIST.

[1]For example, fire departments may wish to have an indoor localization capability in burning buildings with 3 m average accuracy. Similarly, the Federal Communications Commission (FCC) in the US requires the average indoor location accuracy for E-911 calls to be x meters y% of the time.

Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

the R&D community a rich repository of smartphone sensor data, RF signal strength data, and GPS fixes collected based on the guidance provided by ISO/IEC 18305 and (ii) developing a web portal that uses ISO/IEC 18305 to automatically evaluate the performance of indoor localization apps developed based on the data repository and publish the results.

The rest of this paper is organized as follows. Section II presents an overview of related work. We present different aspects of our campaign and the properties of the collected data in Section III. Section IV is on validation of the data we collected. Concluding remarks are given in Section V.

II. Related Work

This section provides an overview of two relevant directions in performance evaluation of indoor localization solutions.

A. Indoor Localization Competitions

Lessons learned during the 2014 IPSN / Microsoft Indoor Localization Competition are reported in [2]. The competition used one scenario to evaluate a broad set of solutions, including RF, magnetic, light, and ultrasound-based systems. EvAAL is another popular series of competitions focused on indoor localization solutions for assisted living [3]. [4] presents the results of an indoor localization competition focused on interference robustness of RF-based localization solutions. Indoor localization competitions are attractive, because they provide the possibility of evaluating and objectively comparing different localization solutions in the same environment and similar conditions. However, such competitions are rare due to their labor, time, and cost intensity. Furthermore, although the evaluation environment is the same, not all solutions can be evaluated at the same time. Hence, the temporal variability of environmental conditions reduces the objectivity of the results.

B. Usage of Raw Data Traces

Use of publicly available data traces is a promising solution for mitigating temporal variability of conditions, which is hardly avoidable in indoor localization competitions. There have been a few efforts to virtualize certain aspects of experimental evaluation of localization solutions. VirTIL testbed [5] is focused on the evaluation of RF range-based indoor localization algorithms by providing range measurements made throughout the evaluation area. The EU EVARILOS project's efforts [6] yielded unprocessed low-level RF data, such as Received Signal Strength Indicator (RSSI) and Time of Flight (ToF) measurements, that are used in a wide range of indoor localization solutions, including fingerprinting, hybrid and proximity-based solutions. This paper builds upon that work and focuses on the extensive set of sensors available in today's smartphones, including not only RF-based ones, but also environmental, position, and motion sensors. Moreover, we significantly expand the scope of the data traces by using several evaluation areas and testing scenarios.

III. Data Collection

This section consists of three parts. First, we describe the environment and the scenarios, selected based on guidance from ISO/IEC 18305, we used for data collection. Second, we provide an overview of the collected data. Third, we provide a synopsis of the collected data.

A. Environment and Scenarios

We selected four buildings for data collection. One was an office building, two were industrial shop and warehouse types of buildings, and the fourth was a subterranean structure. Unfortunately, we did not have access to a high-rise building or a single-family house, as recommended by ISO/IEC 18305 in addition to the above types of buildings. The total space covered by these buildings was about 30,000 m^2.

We instrumented these buildings with more than 900 test points, henceforth called dots, installed on the floors. Each dot is a disk of diameter about 3 cm with its center marked for subsequent surveying. The locations of these dots were precisely determined by a surveying contractor. Therefore, the ground truth locations of the dots are known to NIST.

In an effort to capture the differences in qualities of the sensors and RF circuitry in smartphones, we used four Android[2] smartphones for data collection. These smartphones were LG G4 (LG), Motorola Nexus 6 (NX), OnePlus 2 (OP), and Samsung Galaxy S6 (SG). Since we did not want to repeat each data collection scenario four times, once for each smartphone, we devised a mechanism to collect data with all four phones simultaneously. We used armbands to attach two phones to each arm of the person collecting data, as shown in Figure 1. On each arm, one phone faced forward and the other faced to the side away from the person. We connected four cables in parallel to a push-button switch. The other side of the cables were connected to the audio jacks of the four phones. We used this mechanism to send a signal to the phones to create a timestamp in each of them whenever the person collecting data was on top of a dot.

The data that we collected is described in detail in the next section. We collected two types of data, one for training and the other for testing purposes. The data in each type is timestamped, but the training data is also annotated with the ground truth locations of the dots at which the push-button switch was pressed. The training data would allow the app developers to check how well their apps are performing by comparing the location estimates provided by their apps at time instances when the push-button switch was pressed with the ground truth locations at those time instances. In the case of the test data, we do not provide the ground truth locations at time instances at which the push-button switch was pressed. Instead, we ask each app developer to upload the location estimates provided by the app at those time instances on the PerfLoc web portal so that the portal can evaluate the app's performance. Naturally, a developer would do this step when he/she is convinced that his/her app is as good as it can be.

[2]DISCLAIMER- Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

We collected data using a subset of the 14 Test & Evaluation (T&E) scenarios described in ISO/IEC 18305. We did not use all the scenarios, because some did not apply to our data collection campaign. For example, two scenarios in ISO/IEC 18305 call for multiple people to collect data simultaneously. This would be useful if the localization device has peer-to-peer ranging capability or we wish to test the Medium Access Control (MAC) layer issues and determine how many localization devices can be localized by the system simultaneously. Smartphones do not have a peer-to-peer ranging capability and we are focusing on indoor localization apps that passively listen to RF signals of interest, and hence there are no MAC layer issues. Including the training data, we collected data over 38 T&E scenarios in the four buildings. Each scenario called for the person collecting the data to start at a given location outside a building before entering the building and following a predetermined course while pressing the push-button switch at select dots on the course. The scenarios involved different modes of mobility: 1) walking to a dot and stopping for 3 s before moving to the next dot; 2) walking continuously and without any pause throughout the course; 3) running / walking backwards / sidestepping / crawling part of the course; 4) "transporting" the four phones on a pushcart; 5) using elevators, as opposed to stairs, to change floors; 6) leaving the building a few times during a scenario and then reentering through the same door or another.

B. Types of Data Collected

For each scenario in each building we collected six types of data on each smartphone, namely, Wi-Fi, Cellular, GPS, Dots, Sensors, and Metadata, as described below. This data is stored as one or more Google Protocol Buffer Messages [7] in a separate file for each data type

The collected data is composed of:

1) Wi-Fi data: Signal strengths measured from Wi-Fi access points (APs) in range and other information provided by the APs operating at 2.4 and 5 GHz channels. The Wi-Fi scans were performed back-to-back. The frequency of the scans depended on the Wi-Fi chip's firmware and the Wi-Fi driver implementation on a given smartphone. Since these implementations are provided by the manufacturers, the duration of a Wi-Fi scan varied from device to device. Android platform's Wi-Fi connectivity API supports passive scanning only. Scan results were saved as a single protocol buffer message in repeated fields for each AP.

2) Cellular data: Identity information and signal strengths measured from cellular towers in range. The Android operating system continuously scans for cellular signals and a scan cannot be triggered by a user-level application. Therefore, our application periodically requests all observed cell information from the operating system at a 1 Hz update rate and saves the detected cellular tower related information into the file system as repeated fields of the relevant protocol buffer message.

3) GPS data: Detected geophysical position using GPS. During the data collection campaign, whenever the smartphone got a location fix, the GPS information along with the accuracy of that fix was saved as a separate protocol buffer message. We used "fine location" information, which is derived directly from the GPS signals. Since the measurements were performed indoors most of the time, it was not possible to receive the GPS signal the whole time. If the device could not derive a location from the GPS signals, then no data was stored.

4) Dots: Timestamps at dots visited during a scenario. To make sense of the collected data, it is important to know where and when the data was collected. For each dot visited during a scenario, we store a dot-index and a timestamp. The dot-indices are simply successive non-negative integers with 0 reserved for the starting point of the scenario, which may not be a dot, and 1 through n used for the n dots visited during the scenario after the starting location. For each dot-index, there is a timestamp that designates the time when that location was visited. For the scenario that required waiting for three seconds at each of the dots 1 through n, we stored two dot-indices and two timestamps. Specifically, for the i^{th} dot, we used dot-indices $2i-1$ and $2i$ for the time instances we reached that dot and we left it, respectively. Each dot is represented in a separate protocol buffer message with its index and timestamp.

5) Sensors: Built-in environmental, position, and motion sensors that are found in smartphones. The Android platform defines and supports a number of sensors, which are either hardware-based or software-based. Hardware-based sensors, such as accelerometer, magnetometer, or light, are physical components in the device. Software-based sensors, such as linear acceleration and gravity, are virtual sensors that derive their values from one or more hardware-based sensors.

The Android platform does not specify a standard sensor configuration. Hence, the manufacturers can choose any set of sensors to install in their devices. In our measurements we collected data from all of the sensors that were available on any given device. The list of Android-defined sensors, their descriptions, and whether they are hardware- or software-based are given in Table I. Some sensors are "uncalibrated" versions of others with the same name. These sensors provide additional raw values along with some bias. These types of sensors can be useful when an application conducts its own sensor fusion. More information on uncalibrated sensors can be found in [9]. A smartphone may contain device-specific, non-standard sensors that are not defined in the above list. The information collected by any of the sensors, including the non-standard ones, is saved as protocol buffer messages. Some devices also implement software-based one-shot composite sensors, such as glance gesture, pick up gesture, significant motion, and wake up gesture. These sensors, also called trigger sensors, are used for end-user convenience, such as to briefly turn the screen on or to mimic press of the power button. Since these convenience features do not provide any clear benefits to LTSs, we did not implement support for them, but they may appear in the collected data if they are triggered. Android's sensor framework uses a standard 3-axis coordinate system depicted in Figure 2. It is important to note that the axes of this coordinate system do not change or swap if the orientation of the device changes. For example, the z-axis remains pointing

Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

TABLE I: List of Android sensors [8]

Sensor	Description	Type	Sensor	Description	Type
AMBIENT_ TEMPERATURE	Ambient air temperature	Hardware	LIGHT	Luminance	Hardware
PRESSURE	Ambient air pressure	Hardware	RELATIVE_ HUMIDITY	Ambient relative humidity	Hardware
ACCELEROMETER	Acceleration force along the x,y,z axes	Hardware	GRAVITY	Force of gravity along the x,y,z axes	Hardware or Software
GYROSCOPE	Rate of rotation around the x,y,z axes	Hardware	GYROSCOPE_ UNCALIBRATED	Rate of rotation (without drift compensation) around the x,y,z axes	Software
LINEAR_ ACCELERATION	Acceleration force along the x,y,z axes (excluding gravity)	Hardware or Software	ROTATION_VECTOR	Rotation vector component along the x,y,z axes and Scalar component of the rotation vector	Hardware or Software
STEP_COUNTER	Number of steps taken by the user since the last reboot when the sensor was activated.	Software	GAME_ROTATION_ VECTOR	Rotation vector component along the x,y,z axes	Software
GEOMAGNETIC_ ROTATION_VECTOR	Rotation vector component along the x,y,z axes	Software	MAGNETIC_FIELD	Geomagnetic field strength along the x,y,z axes.	Hardware
MAGNETIC_FIELD_ UNCALIBRATED	Geomagnetic field strength (without hard iron calibration) and Iron bias estimations along the x,y,z axes	Software	ORIENTATION	Azimuth, Pitch and Roll	Software
PROXIMITY	Distance from object	Hardware			

outwards of the screen even when the person collecting data is crawling on the floor instead of walking. Each sensor component has a different reporting frequency. When the app registers to a sensor at the fastest rate, the reporting frequency can be as fast as 250 Hz. Unfortunately, asking for the fastest possible rate sometimes causes individual sensors to be deprioritized and starved. For example, the Samsung smartphone reported far fewer readings from the pressure sensor when all the sensors were registered at the fastest rate. To avoid loss of important data, we programmatically registered to the sensors at a 100 Hz rate, which is commonly selected in research on human motion recognition using motion sensors like accelerometers. This gave all sensors enough time to report their values. Temperature, light, proximity, and humidity sensors, and the step counter generate values only if their last measured values have changed.

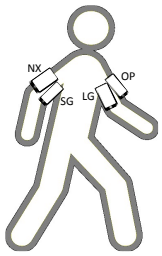


Figure 1: Positioning of the devices on the test subject's body

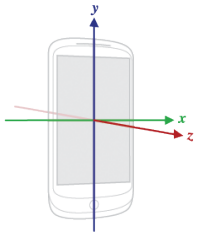


Figure 2: Device-relative coordinate system used by the Sensor API [9]

6) Metadata: The metadata includes the building ID, scenario ID, and measurement device's manufacturer, model, ID, brand, etc. It also includes the initial barometer value, if the smartphone has one, by averaging the first 5 pressure sensor measurements at the beginning of a data collection scenario. Knowing the initial pressure value can help the indoor localization app to detect elevation changes due to movement from one floor of the building to another. The list of sensors that a particular smartphone is equipped with along with their properties is also included in the metadata. For data collection in each building one protocol buffer message for the metadata in each smartphone is generated.

C. A Synopsis of the Collected Data

The number of scenarios used (the time it took to collect the data) in Buildings 1-4 were 11 (∼4.1 hours), 10 (∼5.3 hours), 9 (∼4.4 hours), and 8 (∼1.8 hours), respectively, resulting in a total of 38 scenarios and roughly 15.6 hours of raw data traces for each device. Tables II-V provide an overview of Wi-Fi, cellular, GPS, and sensor data traces, respectively, in two scenarios for each of the four buildings.

The reader can extract similar information for the remaining 30 scenarios by downloading data traces from the PerfLoc web portal. Certain observations can be made about the data in Table II. First, the average durations of a Wi-Fi scan over the 8 scenarios for the LG G4, Motorola Nexus 6, OnePlus 2, and Samsung Galaxy S6 devices (ordered alphabetically) are 3238, 916, 2351, and 3351 ms, respectively. Second, the LG G4 device detects far fewer Wi-Fi APs than the other three devices that detect similar numbers of APs. This is due to the fact that the default configuration of the Wi-Fi scanning procedure in the LG G4 device does not report hidden Wi-Fi APs, i.e. those without the SSID parameter defined. Indeed, if we filter the hidden APs for the other devices for Scenario 1 in Building 1, the number of unique Wi-Fi APs detected would be 56 (unchanged), 61, 59, and 61, respectively. The same observation is made when hidden APs are filtered out in the other scenarios. Additional small variability in the number of detected APs is due to slight differences in the locations and movement patterns of devices, as worn by the person collecting the data, which results in different levels of signal attenuation and shadowing. This is consistent with a previous report of RSSI not being a fully stable feature of Wi-Fi signals [10].

As can be seen from Table III, the number of cellular scans is roughly the same as the duration of a scenario in seconds, except for Building 1, where cellular coverage was far worse than the other three buildings. This is consistent with the 1 Hz cellular scan rate mentioned in Subsection III.B. Note that we did not log empty cellular readings, which explains the slight differences in the duration of scenarios. In addition, the LG G4 and Samsung Galaxy S6 devices had SIM cards and subscriptions to CDMA (C) and LTE (L) data services. Consequently, the data traces for these devices consist of CDMA and LTE signal strengths, although WCDMA (W)

Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

TABLE II: Selected features of the Wi-Fi data traces

Building/ Scenario	Number of Wi-Fi scans				Number of unique Wi-Fi APs				Duration [s]			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG
B1/S1	245	953	356	248	56	150	143	158	836	843	835	837
B1/S2	370	1397	524	358	24	70	70	67	1212	1219	1212	1212
B2/S1	741	2442	983	707	237	516	611	625	2349	2348	2352	2347
B2/S2	792	2667	1057	759	222	515	574	630	2518	2527	2519	2522
B3/S1	696	2545	985	677	70	166	178	190	2288	2298	2288	2291
B3/S2	460	1710	650	447	64	144	141	164	1515	1514	1518	1511
B4/S1	360	1283	496	347	39	104	118	105	1158	1164	1156	1159
B4/S2	207	731	285	198	30	80	77	96	660	661	664	658

TABLE III: Selected features of the cellular data traces

Building/ Scenario	Number of cellular scans				Number of unique cellular BSs				Duration [s]			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG
B1/S1	130	132	166	119	C:3,G:1,L:6	G:1,W:7	G:1,W:6	C:2,L:6	623	136	637	121
B1/S2	47	199	208	201	C:2,L:5	G:1,W:5	G:1,W:7	C:3,G:1,L:5	1216	1221	1215	1219
B2/S1	2311	2273	2327	2301	C:5	W:5	W:6	C:9,L:9	2352	2349	2354	2352
B2/S2	2455	2443	2491	2468	C:6	G:1,W:7	G:1,W:7	C:9,L:8	2523	2528	2522	2528
B3/S1	2108	2215	2263	2043	C:6,W:2	G:1,W:10	G:1,W:13	C:7,L:13,W:1	2293	2299	2292	2294
B3/S2	1413	1469	1505	1463	C:8,G:1,W:3	W:2	W:3	C:7,L:6	1518	1515	1521	1517
B4/S1	1149	1121	1146	1141	C:4,L:22	W:12	W:11	C:5,L:13	1161	1166	1159	1163
B4/S2	658	641	657	651	C:4,L:13	W:4	G:1,W:7	C:4,L:10	664	662	666	663

and GSM (G) readings occur also. For the other two devices, the stored signal strength data are of the WCDMA and GSM types only.

TABLE IV: Selected features of the GPS data traces

Building/ Scenario	Number of GPS readings				Duration [s]			
	LG	NX	OP	SG	LG	NX	OP	SG
B1 / S1	20	12	25	10	18	85	23	13
B1 / S2	54	4	13	7	76	2	11	7
B2 / S1	619	65	179	242	1784	2076	2090	2335
B2 / S2	669	217	269	350	1915	1804	1843	2175
B3 / S1	585	187	180	345	2197	2077	928	2133
B3 / S2	33	76	219	140	31	191	228	195
B4 / S1	890	583	889	790	1150	841	1111	1158
B4 / S2	568	448	561	434	599	520	601	548

Table IV shows that the GPS signal was not always available during a scenario. Just as in the case of the cellular signal, the availability of the GPS signal was far sparser in Building 1 compared to the other 3 buildings. The data shows significant variability across the four devices in both the duration of time the GPS signal was available and the number of GPS readings. In addition, this variability appears to be random, and hence one cannot say that the GPS receiver in any device was better than those in the other devices.

Finally, the numbers of readings from select sensors deemed to be more useful for indoor localization purposes are provided in Table V. These sensors are light, pressure, accelerometer / gyroscope, magnetometer, and step detector (see [2] and the references therein). We have grouped the accelerometer and gyroscope together, because the numbers of readings from these sensors were always within one of each other. Note that the numbers of magnetometer readings for the OnePlus 2 and Samsung Galaxy S6 devices are always within one of the numbers of readings from the respective accelerometer / gyroscope. In addition, these numbers and the numbers of accelerometer / gyroscope readings for the LG G4 and Motoroal Nexus 6 devices are roughly hundred times the duration of the respective scenarios in seconds, which is consistent with the 100 Hz sensor sampling rate mentioned in Subsection III.B. The variation in the numbers of readings from a given sensor across the devices is primarily due to the differences in susceptibility thresholds of the sensors. The numbers of readings from the light and pressure sensors and the step detector are far fewer. Note that OnePlus 2 does not have a pressure sensor. Peculiarly, this device did not detect any steps in Scenarios 1 and 2 in Building 4!

IV. Data Validation

We took certain measures prior to the start of our extensive data collection campaign to ensure the data we were getting from the phones was sound and made sense. For example, we walked in a building following a predetermined course while holding a phone in hand horizontally and pointing to the direction of movement. We accelerated and decelerated in certain places, made turns, went up and down the stairs, violently shook the phone at certain times, turned the lights off and on in one corridor, and pressed the push-button switch whenever we went over a dot. We knew on which accelerometer and gyroscope axes to expect activity at what times, and we verified that that was indeed the case. The data from the pressure and light sensors made sense, and the number of timestamps generated by the switch was the same as the number of dots we visited. Due to lack of space, we do not present the details of that experiment.

However, the problem of validating the extensive data we collected with multiple devices is a lot more involved than that. A major challenge is to figure out what questions to ask and which tests to perform before even deciding whether the questions are adequately answered or the data passed the tests. We suspect more can be done than what is presented next in this section, but at least we address two issues. One is how periodic different sensor and RF signal strength data is. We did not have any influence on how the OS delivers the sensor data to the application level, and we could only suggest a delivery rate without any guarantee on the inter-sample time of individual sensors. This resulted in potentially irregularly sampled time series. This issue is important, because dealing with periodic data would be much easier for the researchers who would use our data to develop indoor localization apps than dealing with asynchronous data. Another question is

Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

TABLE V: Number of sensor readings for selected sensors

Building/ Scenario	Light				Pressure				Accelerometer / gyroscope				Magnetometer				Step detector			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG
B1/S1	3691	2285	3215	4708	64252	27276	0	4708	85294	83911	83284	84748	77467	63648	83285	84748	996	979	949	911
B1/S2	4760	2302	4198	6812	93371	39362	0	6812	123678	121265	120811	122620	112771	91821	120810	122620	788	712	714	196
B2/S1	14673	9318	14491	13135	181186	75854	0	13135	239020	233363	233834	236483	217952	176991	233835	236483	1915	2019	1660	1750
B2/S2	14957	9358	15277	14117	194411	81388	0	14117	256346	251116	250536	254168	234384	189892	250536	254168	1451	1349	1262	802
B3/S1	15515	9780	15966	12823	175718	74232	0	12823	232945	228348	227674	230841	211903	173211	227674	230841	2722	2841	2544	2711
B3/S2	9463	4826	9736	8478	116748	48902	0	8478	154316	150562	151077	152636	140812	114088	151078	152636	1030	1004	974	506
B4/S1	8493	4963	8654	6503	89041	37620	0	6503	118022	115789	115266	117063	107381	87780	115266	117063	1231	1155	0	988
B4/S2	4225	2473	5105	3710	51027	21362	0	3710	67613	65751	66230	66772	61696	49848	66229	66772	527	529	0	330

TABLE VI: Summary of inter-sample times [mean value±standard deviation]

Building/ Scenario	Wi-Fi [s]				Accelerometer / Gyroscope [ms]				Magnetometer [ms]		Cellular [s]			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	LG	NX	OP	SG
B1 / S1	3.3±0.2	0.9±0.0	2.4±0.5	3.4±0.1	9.8±3.7	10.1±6.1	10.1±3.8	9.9±4.5	10.8±4.6	13.3±6.2	4.8±43.2	1.0±0.0	3.9±36.5	1.0±0.0
B1 / S2	3.3±0.1	0.9±0.0	2.3±0.0	3.4±0.1	9.8±3.5	10.1±5.8	10.1±3.7	9.9±4.6	10.8±4.4	13.3±5.9	26.5±170.6	6.1±59.4	5.9±57.0	6.1±49.0
B2 / S1	3.2±0.1	1.0±0.1	2.4±0.4	3.3±0.1	9.8±3.6	10.1±7.3	10.1±4.2	9.9±5.4	10.8±4.5	13.3±7.7	1.0±0.1	1.0±0.0	1.0±0.0	1.0±0.0
B2 / S2	3.2±0.1	0.9±0.1	2.4±0.4	3.3±0.1	9.8±3.8	10.1±7.4	10.1±4.1	9.9±5.6	10.8±4.7	13.2±7.9	1.0±0.2	1.0±0.0	1.0±0.0	1.0±0.0
B3 / S1	3.3±0.1	0.9±0.1	2.3±0.0	3.4±0.1	9.8±4.1	10.1±6.1	10.1±4.0	9.9±5.1	10.8±4.9	13.3±6.2	1.1±1.3	1.0±0.0	1.0±0.0	1.1±1.6
B3 / S2	3.3±0.1	0.9±0.0	2.3±0.4	3.4±0.0	9.8±4.8	10.1±5.7	10.1±4.0	9.9±4.9	10.8±5.6	13.3±5.8	1.1±1.1	1.0±0.0	1.0±0.0	1.0±0.0
B4 / S1	3.2±0.0	0.9±0.0	2.3±0.0	3.4±0.0	9.8±3.6	10.1±6.3	10.1±4.1	9.9±4.6	10.8±4.5	13.3±6.1	1.0±0.0	1.0±0.0	1.0±0.0	1.0±0.0
B4 / S2	3.2±0.0	0.9±0.0	2.3±0.0	3.3±0.0	9.8±3.4	10.1±5.9	10.1±4.1	9.9±4.6	10.8±4.3	13.3±6.2	1.0±0.0	1.0±0.0	1.0±0.0	1.0±0.0

TABLE VII: Summary of Spearman's correlation coefficient ρ for selected sensor types

Building/ Scenario	Wi-Fi						Accelerometer					
	LG-NX	LG-OP	LG-SG	NX-OP	NX-SG	OP-SG	LG-NX	LG-OP	LG-SG	NX-OP	NX-SG	OP-SG
B1 / S1	0.876	0.883	0.884	0.834	0.904	0.835	0.854	0.875	0.845	0.834	0.836	0.804
B1 / S2	0.879	0.876	0.874	0.834	0.902	0.884	0.786	0.832	0.821	0.883	0.856	0.832
B2 / S1	0.856	0.885	0.898	0.874	0.901	0.834	0.864	0.823	0.882	0.843	0.884	0.832
B2 / S2	0.867	0.902	0.829	0.871	0.890	0.897	0.864	0.812	0.845	0.812	0.850	0.842
B3 / S1	0.902	0.875	0.908	0.843	0.876	0.865	0.874	0.831	0.810	0.831	0.845	0.791
B3 / S2	0.896	0.880	0.879	0.800	0.891	0.841	0.831	0.829	0.871	0.811	0.851	0.803
B4 / S1	0.852	0.847	0.902	0.886	0.904	0.906	0.811	0.832	0.814	0.824	0.823	0.841
B4 / S2	0.863	0.912	0.901	0.876	0.901	0.877	0.822	0.851	0.812	0.767	0.855	0.842

whether the data collected with the four devices is "similar", with the caveat that for certain sensors one should not expect similarity due to the differences in where the devices were worn on the person who collected the data and the devices' movements. If we knew the ground truth for all the sensor and RF signal strength data, then we could check whether the data collected with any given device was sufficiently close to the ground truth. In the absence of such ground truth data, all we can do is to decide, for each sensor or RF signal strength, which devices generated similar data. Such tests would reveal, for example, whether three devices generated similar data but the fourth one was stuck at a value due to sensor malfunction or it generated unrealistic erratic or random data because it was not properly calibrated. There can also be programmatic errors in the code that collects and stores data. Even if none of these problems exists, the data collected by the four devices may not be as similar as desired due to the quality of the components used in the phones. For example, the price range for Inertial Measurement Units (IMUs) is from tens of dollars to thousands of dollars. While the more expensive IMUs make more precise measurements, there is much greater variation and uncertainty in the performance of low-end IMUs. These issues would certainly affect the performance of the indoor localization apps to be developed.

Table VI presents the statistics of inter-sample times for select sensor and RF signal strength data that were sampled at vastly different frequencies for various devices and two scenarios in each of the four buildings. Sensors such as light or step detector, which provide readings only when an event occurs, were not included in the table. We observed greater variation in Wi-Fi inter-sample times for the OnePlus 2 device, mostly due to outliers with scan durations of roughly 12 seconds that consistently occur for this device in all scenarios. Figure 3 depicts the relationship between Wi-Fi inter-sample times and the number of APs detected for Scenario 1 in Building 1. There is correlation between the number of APs detected and inter-sample time for the Nexus 6 and OnePlus 2 devices. Similar patterns are observed for other buildings and scenarios. Note that OnePlus 2 outliers are not shown in the figure.

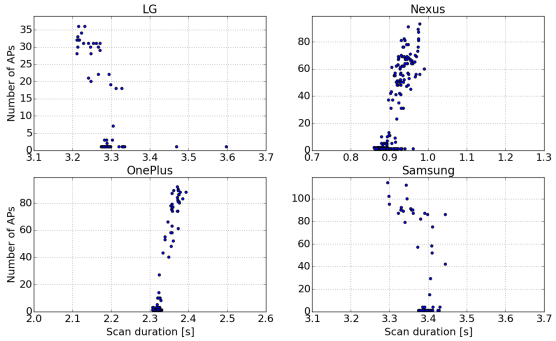


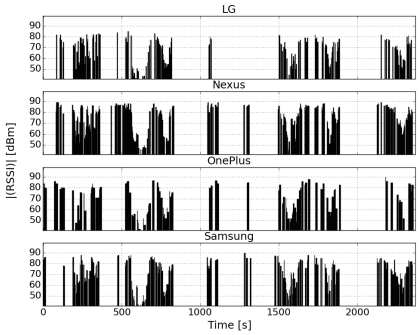
Figure 3: Number of detected Wi-Fi APs vs. scan duration

Table VI shows significant variation in inter-sample times for accelerometer / gyroscope and magnetometers. There are two reasons for this behavior. First, the intended sampling frequency of 100 Hz is not a hard requirement for devices' Operating Systems (OSs). Hence, certain variation in inter-sample time is tolerated by the OS. Second, the time to log the sensor readings, and hence the inter-sample time, is larger

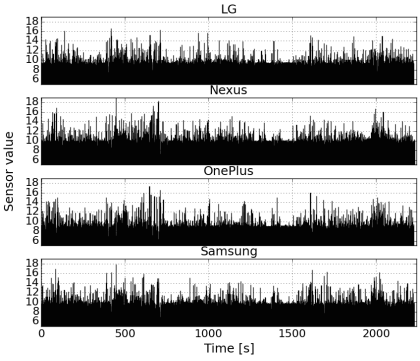
Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

when many sensors report readings. Conversely, those times are shorter when fewer sensors need to report readings. The OS balances these times so that the 100 Hz sampling frequency is roughly realized. Anyone who uses our data to develop an indoor localization app should regard the sensor readings as averages over variable-length intervals of time. Except for most scenarios in Building 1 and some scenarios in Building 3, where cellular coverage was non-existent or weak, the inter-sample times for cellular readings show little deviation from the 1 s target.

To study the similarity of readings from sensors of the same type, e.g. pressure sensor readings from the four devices, we compute Spearman's correlation coefficient ρ and corresponding p-values for all six possible pairs of devices. Due to lack of space, we show only the results for Wi-Fi signal strength and accelerometer data in Table VII. As visible in the table, all correlation coefficients are fairly high, i.e. close to the maximum value of 1. That means the data from pairs of devices are strongly correlated. The p-value indicates the significance of the Spearman's correlation coefficient. We did not explicitly report the p-values in the table, since they tend to zero, meaning that the null hypothesis of the combination of datasets not being correlated is negligibly small.



(a) Wi-Fi data trace



(b) Accelerometer data trace

Figure 4: Example data traces for all four devices

We had to deal with the problems of having slightly different start and end times for data collection in a scenario and different numbers of data samples for a given sensor or RF signal strength from various devices before we could compute any correlation coefficient. For both the RF signal strength and accelerometer readings we solved this problem by merging the sampling times from all phones to a full set of time instances. We then interpolated the missing values and used these newly generated time series for calculating correlation coefficients. Additionally, for the accelerometer, in order to reduce the impact of transients spikes, before the interpolation step, we averaged the data over a sliding window with duration of 100 ms and a sliding step of 20 ms. Example Wi-Fi and accelerometer data traces are given in Figure 4 as a pictorial evidence of data similarity. In the accelerometer case, the depicted data is the l_2-norm of the x, y, and z values of each reading. In the Wi-Fi case, we focus on the AP with the largest number of observations in a particular scenario, since such APs are, in contract to APs with less observations, more relevant for the majority of localization solutions that leverage Wi-Fi readings (e.g. [11], [12]).

V. Conclusions

The data we have collected in this project and are making available to the R&D community is truly unique in the world. We doubt many organizations would have the resources to instrument four large buildings, covering about 30,000 m^2 of space, with 900+ test points, have the locations of the test points professionally surveyed, and spend about 200 man-hours on data collection using four Android phones after months of preparation. We presented some analysis that speaks to the validity of the collected data. Researchers across the world will be able to use our annotated data to develop Android indoor localization apps. Our future plans include development and launch of a web portal for comprehensive performance evaluation of indoor localization apps based on the ISO/IEC 18305 standard. In addition, lessons learned from this and other planned indoor localization "system testing" activities will result in improvements to ISO/IEC 18305. Future part(s) of the paper will describe the performance evaluation web portal and present the results of evaluating a number of smartphone indoor localization apps through the portal.

References

[1] ISO/IEC DIS 18305, *Test and evaluation of localization and tracking systems*, (2015). [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:18305:dis:ed-1:v1:en.

[2] D. Lymberopoulos *et al.*, "A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned," in *Proc. IPSN*, 2015, pp. 178–189.

[3] P. Barsocchi *et al.*, "Evaluating ambient assisted living solutions: the localization competition," *IEEE Pervasive Computing*, vol. 12, pp. 72–79, Oct.-Dec. 2013.

[4] F. Lemic *et al.*, "Experimental evaluation of RF-based indoor localization algorithms under RF interference," in *Proc. ICL-GNSS*, 2015, pp. 1–8.

[5] S. Schmitt *et al.*, "A virtual indoor localization testbed for wireless sensor networks," in *Proc. IEEE SECON*, 2013, pp. 239–241.

[6] F. Lemic *et al.*, "Web-based platform for evaluation of RF-based indoor localization algorithms," in *Proc. IEEE ICCW*, 2015, pp. 834–840.

[7] Google, *Protocol buffers*, (2015). [Online]. Available: https://developers.google.com/protocol-buffers/.

[8] ——, *Sensors overview*, (2016). [Online]. Available: http://developer.android.com/guide/topics/sensors/sensors_overview.html.

[9] ——, *Sensors types*, (2016). [Online]. Available: https://source.android.com/devices/sensors/sensor-types.html.

[10] G. Lui *et al.*, "Differences in RSSI readings made by different Wi-Fi chipsets: a limitation of WLAN localization," in *Proc. ICL-GNSS)*, 2011, pp. 53–57.

[11] M. Azizyan *et al.*, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proc. ACM MobiCom*, 2009, pp. 261–272.

[12] F. Lemic *et al.*, "Experimental decomposition of the performance of fingerprinting-based localization algorithms," in *Proc. IPIN*, 2014, pp. 355–364.

Moayeri, Nader; Ergin, Mustafa; Lemic, Filip; Handziski, Vlado; Wolisz, Adam.
"PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps."
Paper presented at 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain. September 4, 2016 - September 8, 2016.

Organizational Practices in Cryptographic Development and Testing

Julie M. Haney, Simson L. Garfinkel, and Mary F. Theofanos
National Institute of Standards and Technology
Gaithersburg, MD, USA
{julie.haney, simson.garfinkel, mary.theofanos}@nist.gov

***Abstract*—Organizations developing cryptographic products face significant challenges, including usability and human factors, that may result in decreased security, increased development time, and missed opportunities to use the technology to its fullest potential. To better identify these challenges, we explored cryptographic development and testing practices by conducting a web-based survey of 121 individuals representing organizations involved in the development of products that include cryptography. We found that participants used cryptography for a wide range of purposes, with most relying on generally accepted, standards-based implementations as guides. However, many also developed their own implementations and drew on non-standards based resources to inform their development and testing processes. Our results also highlight challenges that incorporating cryptography within products creates within organizations, including the recruitment and management of talent, the product lifecycle, and the ability to explain the security value of products to customers. We conclude by discussing implications of these findings and opportunities for future research.**

***Index Terms*—Cryptography, Usability, Cryptographic standards, Developers**

I. Introduction

As a community of security researchers and practitioners, we are increasingly aware of the pervasive impact that usability and human factors have on the security of information systems. This is especially true in the case of cryptographic development resources (e.g. APIs, standards, and tools) where usability is notoriously poor and has long been regarded as a barrier to development [1], [2]. Cryptographic testing resources fare no better. Certification and testing programs such as the National Institute of Standards and Technology (NIST) Cryptographic Algorithm Validation Program (CAVP) [3] and the National Information Assurance Partnership (NIAP) Common Criteria Protection Profiles [4], [5] currently fail to address usability concerns. Church et al. [6] made a similar observation in 2008, noting that certification procedures that do not consider usability frequently make "unrealistic assumptions as to what the user is capable of" and, as a result, produce a "false perception of what is being assured."

Before making recommendations for increased usability in cryptographic development and testing resources, we must first understand our target population: the organizations and developers that use these resources. Therefore, our research explores the practices and challenges of organizations that are developing products that use cryptography. Our hope is to use this new understanding to improve cryptographic tools and inform greater usability of cryptographic resources.

In this paper, we present the results of a survey of 121 individuals working in organizations that implement cryptography in their products. The survey was guided by the following research questions:

- RQ1: What sources and resources are used for cryptographic implementations?
- RQ2: What cryptographic test and evaluation approaches are employed?
- RQ3: What factors are important to organizations when evaluating the quality of a cryptographic implementation?
- RQ4: How are cryptographic standards used when developing products that implement cryptography?
- RQ5: What challenges do cryptographic developers and their organizations face that are greater than or different from those with non-cryptographic products?

Our research has several contributions. We believe we are the first to ask broader questions to characterize the cryptographic practices and types of resources and standards used by cryptographic developers within *organizations*. Our survey sample differs from many of the cryptography developer research studies to date in that our participant pool consists of professionals who represent organizations and work in cryptographic development primarily on a full-time basis vs. the students or part-time app developers in other studies, for example [7]. This research also offers insights into the challenges that cryptographic implementations introduce into organizational practices from the conceptualization of the product; the assembling of the product team; the design, implementation and testing of the product; and finally to the marketing, sale and end-user support. We believe that this is also the first study to attempt to quantify and rank

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

factors that organizations consider when evaluating the quality of a cryptographic implementation.

II. Related Work

We do not discuss the extensive literature regarding end-user usability of cryptographic products because our focus is on the usability issues of creating and testing those products, not their use.

Decades of research into cryptographic primitives and the development of cryptographic libraries has created a plethora of choices for developers wishing to integrate cryptography into their products. Many security vulnerabilities have resulted from the inability of developers to successfully navigate these choices and securely assemble the cryptographic components into an application. Gutmann [8] observed in 2002 that the powerful functionality of cryptographic libraries frequently resulted in the introduction of security bugs by software developers who did not understand the implications of their choices. In 2004, Nguyen [2] showed that even open-source implementations that are under public scrutiny have cryptographic flaws. Eight years later, Fahl et al. [9] analyzed 13,500 free apps downloaded from the Google Play Store and found that 8% were susceptible to man-in-the-middle attacks, including apps from major financial institutions and established internet providers. Using static analysis, Egele et al. [10] found that 88% of 11,748 examined applications contained a significant error in their use of a cryptographic Application Programmer Interface (API).

Acar et al. [7] performed a systematic analysis of how information resources available to Android app developers impacted code security. They found that participants who were only allowed to use Stack Overflow produced code that was significantly less secure—but more functional—than those who used only the official Android documentation. Nadi et al. [11] analyzed the top Stack Overflow posts on cryptography and, correlating with data from GitHub, concluded that cryptographic APIs were too complex to use reliably, required understanding of the underlying API's implementation, and were at the wrong abstraction level to allow developers to perform common cryptographic tasks.

So what is to be done? One approach is to help developers make better use of APIs and other resources by improving usability and guiding developers to make secure choices [1]. To assist developers, Artz et al. [12] created an interactive plugin for the Eclipse integrated development environment (IDE) to assist Java developers in choosing and integrating the appropriate cryptographic algorithms. Crypto-Assistant [13] and Crypto Misuse Analyzer (CMA) [14] performed similar tasks. However, no user evaluations of these systems were performed, so there remains the of question whether these tools actually result in developers writing code that is more secure. Acar et al. [15] explored the usability of several Python cryptographic APIs. They found that clear documentation with easy-to-use code samples and support for common cryptographic tasks (for example, secure key storage) may be just as, or more important than simple interfaces.

Others alternatively have proposed cryptographic APIs sitting atop new libraries. Forler et al. [16] developed libadacrypt, a cryptographic library written in Ada that is designed to be "misuse-resistant." Bernstein et al. [17] created the Networking and Cryptography library (NaCl), a cross-platform cryptographic library that is designed and implemented to "avoid various types of cryptographic disasters suffered by previous cryptographic libraries such as OpenSSL." Although both libadacrpyt and NaCl were designed to be easier for developers to use, to date, these libraries have not been formally tested for developer usability.

III. Methodology

Between June and July 2016, we conducted a twelve-question, web-based survey targeting individuals with experience developing products that include cryptography. Participants were asked questions related to their cryptographic implementation practices and the challenges their organizations face. The survey questions, informed by discussions with cryptography experts in our organization, contained closed-ended, multiple response ("check-all-that-apply") and five-point scale items, as well as open-ended, free-response items.

The study was approved by the NIST Human Subjects Protection Office. Survey responses were collected and recorded without personal or machine identifiers (e.g., IP address) and not linked back to the recruitment emails or participants. Each survey participant was assigned a reference code that was used for all associated data in the study.

We sent email survey invitations to over 800 individuals on government-industry partnership cryptography mailing lists, 68 individuals from 49 unique companies who had spoken at applied cryptographic development conferences, and 89 U.S. organizations from the Worldwide Encryption Products Survey [18]. Participants were not required to complete all survey questions, but only those who substantively answered at least five of the 10 non-demographic questions were considered. A total of 121 responses were included in the final study results.

For the open-ended responses, we performed iterative, inductive coding on the data to identify general themes. Inductive coding is a commonly used qualitative data analysis approach that allows findings to emerge from the text without the restraint of having to use pre-defined, structured categorizations [19], [20].

IV. Survey Findings

Because not all participants answered all questions, we will specify the number of responses for each question included in our results. For example, if 40 out of 119

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

Table I
PARTICIPANT JOB FUNCTIONS

Job Function Category	n=	%[a]
Managerial (e.g. executive, program or department manager)	17	14%
Cryptographer	11	9%
Developer/Software Engineer	17	14%
Researcher/Educator	9	7%
Security Professional (e.g. security architect, security engineer)	10	8%
Technical - Executive (e.g. CTO, Chief Scientist, Technical Director)	12	10%
Technical - Other (e.g. architect, engineer, certifications)	21	17%
Unknown/not specified	24	20%

[a]Note: percentages do not sum to 100% due to rounding.

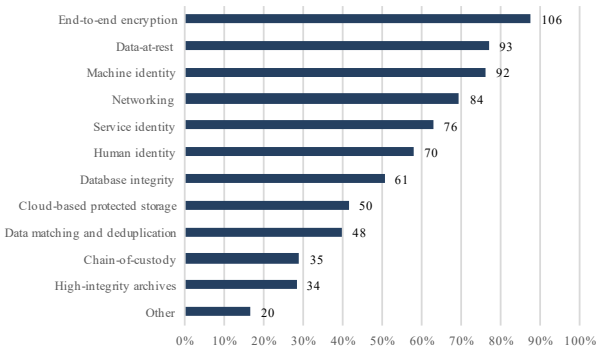


Figure 1. Use of cryptography in products, 769 mentions, 121 respondents.

participants responding to a question selected a particular option, we will indicate that with the notation of "40/119." Also, since there were several multiple responses questions, we also report the number of "mentions" for those questions, in other words, the number of total options that were selected by all respondents for that question. Some response options listed below are abridged from the actual survey text for readability purposes.

Specific participants are referred to with the notation P# where # is between 1 and 121, for example, P23. When providing direct quotes, the participant ID will be followed by a description of their job function (when available). Direct quotes from participants are provided only for those participants who granted permission to quote their responses.

A. Participant Job Functions

Ninety-seven participants specified at least one job function. We categorized reported job functions into high-level categories (see Table 1). Lack of specificity in these open-ended responses (e.g. "engineer") made it impossible to accurately categorize all functions, so a best estimation was made.

Our participant pool showed an obvious skew toward technical roles (80/121), which is appropriate for our research goals. Note that job function categorization into a managerial role may not necessarily mean that the participant lacks technical expertise. The job function responses in our survey also indicated that most of our participants worked on product development or had a role in an organization performing development as their primary job.

B. Product Descriptions

In order to provide some insight into the kind of products represented in the survey, participants were asked, "Where is cryptography used in your products?" Fig. 1 summarizes the responses, showing that end-to-end encryption, data-at-rest, and machine identity were the most commonly specified types of cryptography. Of participants who checked the "Other" option, the two most commonly mentioned categories were software/firmware code integrity (5/20) and hardware-level encryption and security (3/20).

We also asked participants to describe what their product does and how it uses cryptography. It was difficult to categorize these open-ended responses due to the variation in detail and terminology used by participants, so we were not able to do more in-depth statistical analysis of participant responses in relation to product type. Out of the 103 responses, the majority of products were software-based, with only 15 indicating a hardware-based solution. Products were diverse and included operating systems, cryptographic toolkits and libraries, internet of things devices, disk and file encryption, network communication encryption, certificate authorities, authentication software, and key-management solutions, among others.

We also attempted to categorize product descriptions into the number of cryptographic products that each participant represented in the survey. Of the 103 respondents, 40.5% of participants had a single product that uses cryptography, 26.45% had more than one product (but not many products), and 12.4% were from an organization with a large product line. In addition, 1.65% of products were research prototypes or proofs of concept.

C. Cryptographic Implementation Sources and Resources

We asked participants about the sources of the cryptographic implementations used in their products with 82/120 (68.33%) indicating that they use what the hardware, operating system, and standard libraries provide. Almost as many (80/120, 66.67%) said that they develop their own cryptographic implementations and 69/120 (57.5%) selected open-source implementations. Less common were the purchase of commercially available implementations (37/120, 30.83%), requiring customers to purchase specific cryptographic implementations to use with their products

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

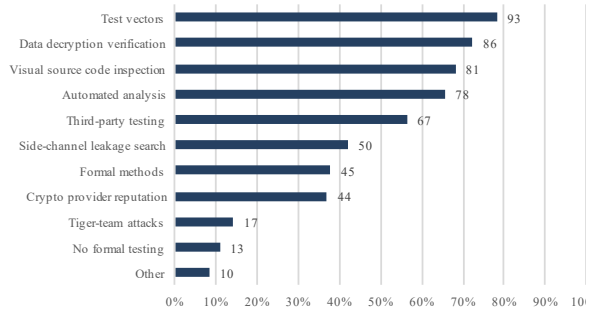


Figure 2. Test and evaluation approaches, 584 mentions, 119 respondents.

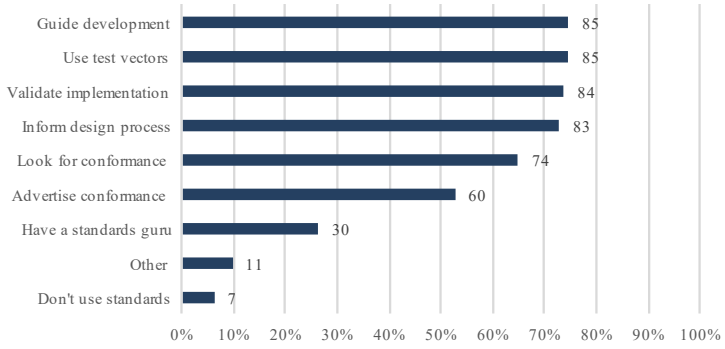


Figure 3. Uses of cryptographic standards, 519 mentions, 114 respondents.

(24/120, 20%), and contracting with others to develop proprietary implementations (15/120, 12.5%).

Participants were also asked what resources they use to help them select or develop cryptographic implementations. Out of 117 responses, an overwhelming majority said that they use national or international standards (109, 93.16%) and industry specifications (94, 80.34%). 66.67% (78) use academic literature, 44.44% (52) use internal (corporate) guidance, 37.61% (44) use web sites, 34.19% (40) use reference books, and 29.06% (34) use proprietary information.

D. Test and Evaluation Approaches

There are a wide variety of approaches that organizations may use for testing, evaluating, and gaining confidence in their cryptographic systems and implementations. To gain greater insight into the incidences of these approaches, we probed participants about their test and evaluation practices. Results are in Figure 2. Ninety-three out of 119 respondents (78.15%) indicated that they test their implementations with specific test vectors (test cases), while 86 (72.27%) said that they verify that the data can be decrypted after it is encrypted. Thirteen (10.92%) said that they do not do formal testing, but would be able to visually observe if data were not being properly encrypted.

Cryptographic certifications are often a customer purchase requirement for cryptographic products. For example, third-party testing laboratories may perform validation testing to the CAVP and NIAP guidelines mentioned earlier. More than half (67/119, 56.3%) specified that they rely on third-party testing for certification or aim for compliance with testing programs. In a related, open-ended question, participants were asked if they employ a third-party testing organization and which one(s) they use. Fifty-three respondents indicated some use of third-party testers, with 21 listing at least one commercial testing lab by name; 20 mentioning NIST, a NIST validation program (e.g. CAVP) or the Federal Information Processing Standard (FIPS) [21]; and 11 mentioning the Common Criteria or NIAP. As discussed in the Limitations section, the participant pool was weighted towards those with an awareness and interest in NIST-related cryptographic resources, which may explain the observed bias towards government-endorsed testing programs.

E. Use of Cryptographic Standards

The survey asked participants if they use cryptographic standards and, if so, how. Approximately 74% (85/114) use standards to guide their development process, with the same number using test vectors from the standards to validate the cryptographic modules. 6.14% (11/114) actually indicated that they don't use standards. See Figure 3 for the response summary.

In two open-ended response questions, participants were also asked which cryptographic standards they use and how they chose those standards. For the types of cryptographic standards, the level of detail provided by respondents varied greatly. Some participants listed specific algorithms (e.g., AES, SHA2, 3DES), others mentioned standards bodies (e.g., IEEE) or policies (FIPS, NIST SPs), and still others indicated the use of standards in generic terms, for example, *"Too many to list"* (P16).

As we were most interested in the sources of standards, we attempted to quantify these open-ended responses by doing frequency counts on mentions of standards authorities. Out of 83 responses, the six most frequently mentioned standards organizations were NIST (including mentions of FIPS) (61), IETF (20), ANSI (14), International Organization for Standards/International Electrotechnical Commission (ISO/IEC) (14), IEEE (13), and Common Criteria/NIAP (12).

Reasons for the choice of cryptographic standards varied. We categorized the 72 open-ended responses, revealing that the most popular reason (31, 43.06%) was market drivers and customer demand, followed by government mandate

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

Table II
IMPORTANCE RANKING OF EVALUATION METRICS FOR CRYPTOGRAPHIC IMPLEMENTATIONS

Rank	Evaluation Metric	Mean	Score 4 or 5	Score 1 or 2
1	General acceptance of algorithms	4.20	81%	10%
2	Runtime protection of secrets	3.92	68%	13%
3	Documentation quality	3.89	66%	12%
4	Code quality	3.88	72%	11%
5	Support for HW that accelerates crypto operation	3.75	63%	15%
6	Throughput	3.73	66%	15%
7	Support for HW that supports crypto key stores	3.69	61%	19%
8	Evaluation by third-party testing organizations	3.68	62%	19%
9	Openness of source code	3.38	48%	27%
10	Implementation language	3.34	49%	25%
11	Implementation size	3.26	42%	27%
12	Evaluation by outside trusted consultants	3.24	46%	27%
13	Power requirements	3.20	43%	28%
14	Blessed by local expert	3.05	36%	35%
15	Popularity in other products	2.92	35%	36%

and recommendations (19, 26.39%), applicability to the domain/problem (10, 13.89%), the quality of the standard (7, 9.72%), interoperability (7, 8.33%), prior familiarity or experience with the standard (5, 6.94%), and internal organizational assessment (4, 5.56%).

F. Quality Evaluation Factors

Participants were asked to individually assess the importance of 15 different metrics, or factors, when evaluating the quality of a cryptographic implementation, with a rating of one being the least important, and five being the most important. 120 participants ranked at least one of the factors. Table 2 shows the resulting ranked order.

There was a notable difference between job function responses. We include the top three and bottom three scored factors for each group in Table 3. General acceptance was ranked highly by all groups except Technical Executives. Popularity was ranked low by all groups. However, there was variation in other factors deemed to be important by each group.

G. Organizational Challenges

In a multi-part, open-ended question, participants were asked about the challenges their organizations face when incorporating cryptography in its products, specifically where cryptography creates challenges that are fundamentally greater or different than other kinds of technology.

1) Recruiting Talent: When asked about recruitment challenges, the majority (42/66, 63.64%) indicated that it was hard to find qualified talent with strong cryptography skills. Three participants commented on the challenge of evaluating cryptographic and secure coding expertise in an interview setting. Since cryptography is a niche skill, these findings are not surprising. However, recruiting employees with just cryptographic knowledge is not the only goal. Nine participants mentioned the need for employing talent who have a combination of skill sets. P22, a software engineering manager at a large company, said, *"It is hard to combine software engineering, math, cryptography, and collaboration skills into one person who would understand the importance of shipping a secure product."*

Table III
MOST AND LEAST IMPORTANT FACTORS PER JOB FUNCTION. () INDICATES A TIE.

Job Function	Top 3 Factors	Bottom 3 Factors
Managerial	Runtime Throughput Acceptance	Openness Popularity (Local expert, Power, Size)
Cryptographer	(Runtime, Third party eval.) (Acceptance, HW crypto key) (Throughput, Documentation)	Popularity Language Openness
Developer/ SW Engineer	Code quality Openness Acceptance	Power Size Popularity
Researcher Educator	Acceptance Runtime Openness	Expert eval. (Popularity, Outside eval.) Language
Security Professional	HW crypto key (Acceptance, Code quality, HW acceleration) (Documentation, Size)	(Popularity, Openness) Language Power
Technical Executive	HW acceleration HW crypto key (Documentation, Throughput)	Expert eval. (Popularity, Outside eval.) Power
Technical Other	Acceptance HW crypto key (Runtime, Documentation)	Popularity Openness Language

To address recruitment problems, eight respondents said they use a hire-and-train strategy. One participant wrote, *"We generally end up hiring smart people and training them on security and cryptography"* (P19, director of development at a security consulting company).

2) Managing Employees and Evaluating Employee Work: Only 17/43 of respondents (39.5%) indicated that there were greater challenges when managing employees. However, nine of these participants indicated that having highly intelligent employees with specialized skills introduces management challenges, especially when managers are not well-versed in cryptography. Furthermore, traditional project management must be adjusted to accommodate the meticulous nature of building quality cryptographic implementations and the characteristics of those working in that field as expressed by one participant: *"Cryptographers tend to drill deep and act conservatively so expectations around how quickly a task will get done need to be adjusted"*

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

(P106, engineering manager at a large software company).

When asked about challenges in evaluating employee work, a little more than half (22/40, 55%) indicated that they face greater challenges. Seven participants commented that an expert or expert source is needed to appropriately evaluate cryptographic work, which is a challenge given the dearth of cryptography experts. Three participants said that it is hard, and perhaps impossible, to know when employee work has reached a truly secure level.

3) Obtaining Appropriate Development Tools: When asked about obtaining appropriate developer tools, 13/40 (32.5%) of responding participants felt that there were greater or different challenges. Of those, the most commonly mentioned challenge (6/40) was in acquiring and having the expertise to use quality bug-finding and testing tools.

4) Maintaining and Transitioning Products: Responses about challenges in product lifecycle maintenance (45 responses) and transition to new products (43 responses) revealed several shared themes. The most common theme, mentioned by 17 participants, related to challenges with maintaining backwards compatibility with older cryptographic algorithms. One participant in particular clearly articulated challenges with the staying power of deprecated cryptography: *"Crypto protocol agility has created a massive problem of zombie algorithms that can't be got rid of. Transitioning to new algorithms is easy. Transitioning from old algorithms is hard"* (P32, random number generator expert at a large company).

Participants further noted a conflict between moving towards state-of-the-art cryptography and maintaining the functionality of legacy applications which are often employed for many years at a time. One participant remarked, "*Old crypto algorithms never die; we'll never get rid of them because somewhere, someone, many years ago, protected data with an old algorithm"* (P95, software engineering manager at a large software company). Another said, "*There is a tension between turning off weak and dangerous algorithms and not breaking users"* (P118, cryptographic library developer).

The update of cryptographic products due to security vulnerabilities or bugs was viewed as challenging by 16 participants. The frequency of security issues in foundational cryptographic implementations and libraries, in particular, requires extra effort as expressed by one participant: *"This is a huge problem. We are constantly needing to evaluate security vulnerability announcements, updating our own products, and then advising customers to update their systems"* (P66, CTO, cybersecurity vendor).

Because of the seriousness of vulnerabilities in cryptographic implementations, several participants commented that updates demand an urgency above and beyond that of other technologies and can be disruptive to the organization. One participant commented, *"Security patches are disruptive and propagate all the way up from the cryptography toolkits through to many of the company products"* (P106, engineering manager at a large software company).

Another theme emerging from the survey data of nine participants was a challenge in keeping abreast of changing standards. One respondent noted, *"Main issue is getting product teams to upgrade to address every changing FIPS requirement and crypto strength requirement"* (P52, software crypto module development engineer). This was an interesting perception given that, in actuality, FIPS standards change infrequently.

A final theme was customer resistance to transitioning to new products, noted by five participants. These participants commented that customers are often hesitant to adopt new cryptographic standards and products, even if they offer greater security. Said one respondent, *"Customers view crypto and crypto libraries as...unchanging components of their systems"* (P65, senior engineering manager at a large vendor).

5) Participating in Product Evaluations: Out of 31 responses, 20 participants indicated that they face greater or different challenges when participating in cryptographic product evaluations. No common themes emerged from these data. However, three participants did note challenges when bringing together multi-disciplinary evaluation teams with different foci and expertise. One participant said:

> *"As usual in the world of security, many product engineers aim at having something working, while cryptographers and security folks focus on the opposite: does it fail where it is supposed to? Secondly, explaining a theoretical cryptography game for an adaptive adversary to a standard software engineer or a program manager simply induces sleep and results in lack of credibility for the cryptographer...[T]he difference in disciplines is just too large."* (P95, software engineering manager)

6) Explaining Products to Potential Customers: Explaining products to customers was viewed as a challenge by 45/48 respondents. Seventeen of these participants said that customers lack security and cryptographic knowledge, making it more difficult to explain a product's functionality or for the customer to properly use the product. One participant commented, *"Many potential customers do not know anything about PKI and our products are all PKI-based so we often end up having to teach them PKI in order to explain our products"* (P19, director of development at a security consulting company). Furthermore, two participants noted that they don't believe that customers even care to understand the underpinnings of the product, with one bluntly stating, *"90% just want a certification check off"* (P22, certifications coordinator for a secure mobility products company).

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

Nine participants commented that it can be difficult to demonstrate the value proposition of their products, and four noted customer concerns about paying for increased security or more robust implementations when what they already have seems to be "good enough." P98, a storage security engineering consultant, commented, *"security adds assurance (insurance) at a cost, but does not provide a way to improve profit."* Another participant wrote about competition with open source implementations: *"In the commercial sector, we have had difficulty convincing potential customers/partners on the value of our products vs. 'free' open source products even with data"* (P39, CEO of a small security solutions company).

Difficulties in convincing customers about the value of cryptographic products may stem from the observation of four participants that many customers view security as an impediment. One participant commented, *"Security is an inconvenience and complicated. Customers need to know a lot to use the products correctly and securely but they don't understand it easily or willingly"* (P106, engineering manager at a large software company).

Five participants remarked on challenges relating to the varying levels of detail they have to provide to their potential customers and what that appropriate level of detail is. Providing detail allows for greater exposure of the strengths of the product, but may also confound the product explanation when customers aren't especially knowledgeable about those details. For example, one participant commented:

> *"Customers don't need to know the details of the crypto unless they're experts, in which case they can understand. The only thing that causes problems is when non-experts insist on having things explained or seeing lab reports and then don't know how to interpret the answers."* (P30, CTO at a cryptographic platforms and services company)

A final theme voiced by four participants was the role of marketing. Three of these participants expressed frustration with marketers that over-inflate claims about their cryptographic products in an attempt to differentiate the security of their product over others. P74, a technologist at a start-up company, stated his issue with product marketing, *"If you don't write hyperbolic claims in slimy marketer-speak, you don't get in the front door even with solid solutions."*

V. Limitations

Our study has several limitations. Since there is no prior research into what is representative of the cryptographic development community, our results may not generalize to the entire population of developers and organizations who implement cryptography. One of the biggest obstacles to generalization is that the sampling frame was heavily biased towards individuals on a government-maintained cryptography mailing list who may be more likely to be aware of and interested in cryptographic standards, especially those published or endorsed by the U.S. Government. Additionally, not all of the individuals on the mailing lists met our criteria of having experience in developing products that use cryptography, as the list may include many who are cryptography researchers, individuals who represent cryptographic certification organizations, and others who have an interest in following the field, but do not develop products. This may account, in part, for the low survey response rate. The sample is also heavily biased towards U.S. organizations who are likely to have different customer sets with different requirements than non-U.S. companies. However, although not generalizable, we believe that our findings are still valuable as they begin to identify gaps and areas for future research.

In addition, individual survey questions were optional, so demographic data for certain items, like job function and organization description, are unknown for some participants. For example, the opportunity to specify a description of the participant's organization was dependent on the participant granting permission to quote responses, but only 35 out of 121 participants granted permission, with 34 providing an organizational description. Responses for open-ended questions about challenges that cryptographic products pose to organizations also had relatively low response rates (between 31 and 66 respondents), perhaps due to the effort required to answer these. Interviews may be more conducive to obtaining this type of data. To this end, we are currently conducting a follow-up interview study to delve deeper into some of the more interesting survey findings.

VI. Discussion and Future Work

Our survey was exploratory in nature, covering a wide swath of organizational practices and challenges in cryptographic development and testing. Throughout the survey, multiple participants specifically cited the difficulty, cost, and scale of cryptographic development. In this section, we discuss some of the more interesting findings that we believe may warrant additional investigation.

A. Usability

Standard libraries, free, and open source software were among the top three sources of cryptographic implementations selected by participants. Unfortunately, these types of implementations are notoriously fraught with usability issues, vulnerabilities, and inadequate constraints that may lead to developers producing insecure code [1], [8], [10]–[12], [22]. The general manager at a cryptography company supported this observation:

> *"In the market, we see a tendency to simply 'borrow' from Open Source implementations without an understanding of the task at hand. As a*

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

result, the crypto deployed is often not deployed [appropriately] which increases the risk of flawed implementations." (P23, general manager at a cryptography company)

Another participant noted, *"Even with standardized libraries, ciphers, etc. it is very easy to implement them poorly and create side channel attacks"* (P96, architect and manager for an open-source security project). This survey supports previous work suggesting that there is a pressing need to improve the usability of cryptographic APIs and associated documentation.

Over 90% of organizations surveyed consult cryptographic standards to help them select or develop cryptographic implementations. Participants also indicated that cryptographic standards are used throughout product design, development, and testing processes. However, today's standards have no emphasis that we are aware of on the usability of programmer APIs, end-user software, or documentation. Furthermore, we have little understanding of the usability of these standards and associated documentation in and of themselves. Given that our survey demonstrates the difficulty of finding skilled professionals to do this type of work, adding usability requirements to and improving the usability of standards might result in reducing the training and domain knowledge that developers require to use cryptography. Further work needs to be done to better understand how the usability of standards can be improved to support both novices and experts.

B. Testing and Evaluating Cryptographic Implementations

Although validated algorithms are in wide use, many implementations of these algorithms are not validated. Validated or not, it is currently unclear how organizations should test their cryptographic software. Over 78% of participants said that they use test vectors, but there are many unanswered questions about these vectors. For example, what are the current shortcomings, if any, of these test vectors? How do we make these test vectors and their accompanying documentation more usable?

Less than half of respondents use more advanced testing approaches such as side-channel leakage search and formal methods. Other studies [12]–[14], [16], [17] contend that there is a lack of quality tools necessary to identify cryptographic misuse within code. However, our study, with a few exceptions, showed an overall lack of concern with respect to obtaining appropriate developer tools when participants were asked about challenges. This might indicate that developers may not understand how to properly test, evaluate, and find vulnerabilities in cryptographic code, and therefore don't realize that the tools they are currently using may be inadequate. How can the research community advance testing tools and ensure they are usable to a wide audience?

There is likewise a need to make the criteria for what constitutes rigorous testing more visible and accessible. This is evidenced by thirteen of our survey respondents who indicated that they do no formal testing, but rather perform a visual inspection of data to determine if their cryptography is operating properly—assuming, perhaps, that plaintext is easily recognizable, and that anything that is not plaintext is encrypted. An obvious problem with this approach is that there are many transformations that render plaintext incomprehensible without providing cryptographic protection, such as Base64 encoding, compression, or encrypting with a constant key.

In addition, we found that organizations often use their own cryptographic implementations, raising questions about the process by which they validate the security of these implementations as well as implications of customers using potentially non-validated, non-certified cryptographic implementations. We believe this is yet another area ripe for additional exploration.

Finally, over half of survey respondents said they use third-party testing, with most of those trying to attain some kind of formal certification to a national or international standard (e.g. FIPS or Common Criteria). However, the cost of testing products may limit participation in these programs. P103, a lead project architect at a cryptographic product development company, commented, "*Cryptographic certifications (e.g., FIPS 140 or Common Criteria) are VERY difficult, expensive, and time consuming - much more so than evaluations for other kinds of products."* These observations point to the need for more in-depth investigation into the perceived benefit vs. cost of certification programs.

C. Evaluation Metrics

As our survey demonstrates, there are many possible ways to evaluate cryptographic implementations. But given the wide variety of choice in cryptographic implementations and the differing opinions of what is important depending on an individual's job function, without clear evaluation metrics, organizations may not choose cryptographic implementations that reflect their goals and priorities. We believe that more work needs to be done to better understand factors affecting implementation decisions and how to design tools to support these decisions.

Trust of standards is a less tangible influence not explicitly asked about in the survey, but worthy of further exploration. Several of the open-ended responses suggested potential issues with trust of standards bodies, particularly the trust of government standards by non-government organizations given concerns in recent years [23], [24]. One survey participant echoed this doubt, saying, "*lost trust in US-based standards [has] become more prominent"* (P95, software engineering manager).

D. Cryptographic Agility

Customer resistance to upgrading to new cryptographic algorithms and products emerged as a challenge to develop-

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

ers wanting to move towards state-of-the-art. For example, many organizations experienced problems when the Message Digest #5 (MD5) cryptographic hash algorithm was deprecated, or during the transition to Security Socket Layer (SSL) 2.0 to SSL 3.0 to TLS. Although we did not specifically ask our respondents what problems they might encounter when cryptographic algorithms are no longer approved for use in an application, the general comments that we received on the difficulties caused by cryptographic agility and the lack of customer understanding of cryptography in general indicates that many will have problems. While the ability to use new cryptographic algorithms may be a boon to developers, algorithmic deprecation clearly causes problems that could be measured quantitatively.

E. Organizational Focus

We believe there is a gap in the literature in exploring organizational, rather than individual, practices in the development of products that use cryptography. Past studies have been useful in highlighting some of the pitfalls of including cryptography within products, but it is unclear how these apply to organizational development and testing. Are organization developer populations more likely to be skilled and use more formal methods for development since the reputation and profit of their company depends on a robust, error-free implementation? Given that 45 respondents in our survey indicated that they face challenges explaining their products to customers, it would also be interesting to explore organizational approaches to communicating the value of cryptography to non-technical audiences.

Our study serves as a first step in understanding these organizational practices and associated challenges. We are currently conducting a follow-up interview study to delve deeper into some of our survey findings, and see the potential for much future research in this area.

References

[1] M. Green and M. Smith, "Developers are not the enemy! The need for usable security APIs," *IEEE Security and Privacy*, vol. 14, no. 5, pp. 40–46, Sept 2016.

[2] P. Nguyen, "Can we trust cryptographic software? Cryptographic flaws in GNU privacy guard v1. 2.3," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Heidelberg: Springer, 2004, pp. 555–570.

[3] National Institute of Standards and Technology, "Cryptographic algorithm validation program (CAVP)," 2017. [Online]. Available: http://csrc.nist.gov/groups/STM/cavp/

[4] D. Leaman, "National Institute of Standards and Technology: NVLAP Common Criteria testing. NISTHB 150-20." 2014. [Online]. Available: https://www.nist.gov/sites/default/files/documents/nvlap/NIST-HB-150-20-2014.pdf

[5] "National Information Assurance Partnership: Approved protection profiles," 2017. [Online]. Available: https://www.niap-ccevs.org/Profile/PP.cfm

[6] L. Church, M. Kreeger, and M. Streets, "Introducing usability to the Common Criteria," in *Proceedings of the 9th International Common Criteria Conference (ICCC)*, Jeju, Korea, 2008.

[7] Y. Acar, M. Backes, S. Fahl, D. Kim, M. L. Mazurek, and C. Stransky, "You get where you're looking for: The impact of information sources on code security," in *2016 IEEE Symposium on Security and Privacy (SP)*, May 2016, pp. 289–305.

[8] P. Gutmann, "Lessons learned in implementing and deploying crypto software." in *Usenix Security Symposium*, 2002, pp. 315–325.

[9] S. Fahl, M. Harbach, T. Muders, L. Baumgärtner, B. Freisleben, and M. Smith, "Why Eve and Mallory love Android: An analysis of Android SSL (in)security," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS '12, 2012, pp. 50–61.

[10] M. Egele, D. Brumley, Y. Fratantonio, and C. Kruegel, "An empirical study of cryptographic misuse in Android applications," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*, 2013, pp. 73–84.

[11] S. Nadi, S. Krüger, M. Mezini, and E. Bodden, "Jumping through hoops: Why do Java developers struggle with cryptography APIs?" in *Proceedings of the 38th International Conference on Software Engineering (ICSE '16)*, 2016, pp. 935–946.

[12] S. Arzt, S. Nadi, K. Ali, E. Bodden, S. Erdweg, and M. Mezini, "Towards secure integration of cryptographic software," in *2015 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (Onward! '15)*, 2015, pp. 1–13.

[13] R. Garcia, J. Thorpe, and M. Martin, "Crypto-Assistant: Towards facilitating developer's encryption of sensitive data," in *12th Annual International Conference on Privacy, Security and Trust (PST '14)*, 2014, pp. 342–346.

[14] S. Shuai, D. Guowei, G. Tao, Y. Tianchang, and S. Chenjie, "Modelling analysis and auto-detection of cryptographic misuse in Android applications," in *2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing*, 2014, pp. 75–80.

[15] Y. Acar, M. Backes, S. Fahl, S. Garfinkel, D. Kim, M. Mazurek, and C. Stransky, "Comparing the usability of cryptographic APIs," in *Proceedings of the 38th IEEE Symposium on Security and Privacy*, 2017.

[16] C. Forler, S. Lucks, and J. Wenzel, "Designing the API for a cryptographic library: A misuse-resistant application programming interface," in *Proceedings of the 17th Ada-Europe International Conference on Reliable Software Technologies (Ada-Europe'12)*, 2012, pp. 75–88.

[17] D. J. Bernstein, T. Lange, and P. Schwabe, "The security impact of a new cryptographic library," in *Proceedings of LatinCrypt 2012*, 2012, pp. 159–176.

[18] B. Schneier, K. Seidel, and S. Vijayakumar, "A worldwide survey of encryption products," Berkman Center, Tech. Rep. Research Publication 2016-2, 2016. [Online]. Available: https://www.schneier.com/academic/paperfiles/worldwide-survey-of-encryption-products.pdf

[19] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory: Strategies for qualitative research.* Transaction Publishers, 2009.

[20] D. R. Thomas, "A general inductive approach for analyzing qualitative evaluation data," *American Journal of Evaluation*, vol. 27, pp. 237–246, Jun. 2006.

[21] National Institute of Standards and Technology, "FIPS Pub 140-2: Security requirements for cryptographic modules," 2001. [Online]. Available: http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf

[22] D. Lazar, H. Chen, X. Wang, and N. Zeldovich, "Why does cryptographic software fail?: A case study and open problems," in *Proceedings of 5th Asia-Pacific Workshop on Systems (APSys '14)*, 2014, pp. 7:1–7:7.

[23] L. Greenemeier, "NSA efforts to evade encryption technology damaged U.S. cryptography standard," Sep. 2013. [Online]. Available: https://www.scientificamerican.com/article/nsa-nist-encryption-scandal/

[24] National Institute of Standards and Technology, "NIST removes cryptography algorithm from random number generator recommendations," Apr. 2014. [Online]. Available: https://www.nist.gov/news-events/news/2014/04/nist-removes-cryptography-algorithm-random-number-generator-recommendations

Haney, Julie; Garfinkel, Simson; Theofanos, Mary.
"Organizational Practices in Cryptographic Development and Testing."
Paper presented at 2017 IEEE conference on communications and network security (CNS), Las Vegas, NV, United States. October 9, 2017 - October 11, 2017.

Photo-Induced Magnetic Field Effects in Single Crystalline Tetracene Field-Effect Transistors

Hyuk-Jae Jang[a,b], Emily G. Bittle[b], David J. Gundlach[b], and Curt A. Richter[b]

[a] *Theiss Research, La Jolla, CA 92037, USA, hjaejang@gmail.com,* [b] *Engineering Physics Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA.*

Development of organic field-effect transistors (OFETs) is considered to be one of the most prominent achievements in advancing organic semiconductor device technology [1], [2]. Several decades of research on OFETs has led to significant advances in the fundamental understanding of charge carrier transport mechanisms in small-molecule organic semiconductors as well as conjugated polymers [1], [2]. However, relative to other organic-based devices such as organic light emitting diodes (OLEDs) [3] and organic photovoltaics (OPVs) [4], OFETs have not been significantly exploited to understand spin related phenomena and their associated magnetic field effects (MFEs) in organic materials even though OFETs can provide a unique scientific platform to study the dynamics of spin interactions such as the spin dependent recombination of excitons and charge transfer pairs [5], [6].

We present an investigation of the change of photo-induced current in single crystalline tetracene field effect transistors as a function of an external magnetic field to elucidate dynamics and interactions of photo-excited spin states in tetracene. Our field effect transistors consist of a heavily doped silicon substrate (gate electrode, n-type 10−3 Ω cm), 200 nm thick thermally grown silicon oxide (SiO_X) layer (gate dielectric), photolithographically defined metal electrodes made of 40 nm thick platinum on 2 nm thick titanium (source and drain contacts), and a tetracene single crystal. The tetracene single crystals were grown by physical vapor transport in a tube oven under argon flow and carefully laminated to the surface of a substrate with electrodes. Completed transistors have channel lengths of 5 μm, 10 μm, and 20 μm. A self-assembled monolayer of octadecyltrichlorosilane (OTS) was used to improve the semiconductor adhesion and to create a hydrophobic surface on the SiO_X to eliminate water [7]. Electrical measurements were carried out in a commercially available probe station at room temperature. The measurements were performed in a nitrogen gas filled environment to minimize possible device degradation by oxidation.

Fig. 1 shows the measured drain-source current (I_{DS}) *versus* drain-source voltage (V_{DS}) with different values of the gate voltage (V_G) of a single crystalline tetracene transistor with a channel length of 10 μm under an illumination intensity of 0.34 kW/m^2. The I_{DS} was much larger under illumination than in darkness (more than twice in magnitude). This increase indicates the existence of photo-induced current in tetracene. When an external in-plane magnetic field was applied, the magnitude of I_{DS} further changed. A decrease in the I_{DS} was observed (up to ≈ 4%) at an applied field of 200 mT under an illumination intensity of 0.34 kW/m^2. However, in darkness, no change in the I_{DS} was detected with an external magnetic field implying that the magnetic field dependence in tetracene is related to photo-induced phenomena. Fig. 2 displays the measured magnetoconductance (MC) at 200 mT as a function of the illumination intensity where MC is defined as $[I_{DS}(B) - I_{DS}(B = 0)] / I_{DS}(B = 0)$. We found that the magnitude of MC increased as the illumination intensity increased. In contrast, a decrease in the magnitude of MC was observed when either the V_G or the V_{DS} was increased. In addition, we found that the magnetic field dependence of MC changes when the angle between the magnetic field axis and the direction of I_{DS} was altered. Theoretical analysis suggests that this angular dependence likely resulted from the change of crystallographic orientation in the magnetic field

Tetracene is known to produce singlet fission processes under illumination, and previous luminescence and fluorescence spectroscopy studies suggested that an external magnetic field disturbs singlet fission as well as triplet fusion processes that also exist in tetracene single crystals [8], [9]. Our study on the MC of illuminated single crystalline tetracene transistors are in agreement with these previous findings. Our analysis suggests that triplet exciton-charge carrier interactions and magnetic

Jang, Hyuk-Jae; Bittle, Emily; Gundlach, David; Richter, Curt; Zhang, Qin.
"Photo-Induced Magnetic Field Effects in Single Crystalline Tetracene Field-Effect Transistors."
Paper presented at International Semiconductor Device Research Symposium (ISDRS 2016), Bethesda, MD, United States. December 7, 2016 - December 9, 2016.

dipolar interactions in triplets play an important role in the observed changes in I_{DS} under an external magnetic field with different illumination intensities, V_{DS}, V_G, and field angles.

References

[1] M. Muccini, "A bright future for organic field-effect transistors," *Nature Mater*. vol. 5, pp. 605-613, 2006.
[2] H. Sirringhaus, "25th anniversary article: organic field-effect transistors: the path beyond amorphous silicon," *Adv. Mater*. vol. 26, pp. 1319-1335, 2014.
[3] H.-J. Jang, S. J. Pookpanratana, A. N. Brigeman, R. J. Kline, J. I. Basham, D. J. Gundlach, C. A. Hacker, O. A. Kirillov, O. D. Jurchescu, and C. A. Richter, "Interface engineering to control magnetic field effects of organic-based devices by using a molecular self-assembled monolayer," *ACS Nano* vol. 8, pp. 7192-7201, 2014.
[4] B. Hu, L. Yan, and M. Shao, "Magnetic-field effects in organic semiconducting materials and devices," *Adv. Mater*. vol. 21, pp. 1500-1516, 2009.
[5] M. Nishioka, Y.-B. Lee, A. M. Goldman, Y. Xia, and C. D. Frisbie, "Negative magnetoresistance of organic field effect transistors," *Appl. Phys. Lett*. vol. 91, pp. 092117-092119, 2007.
[6] T. P. I. Saragi and T. Reichert, "Magnetic-field effects in illuminated tetracene field-effect transistors," *Appl. Phys. Lett*. vol. 100, pp. 073304-073306, 2012.
[7] E. G. Bittle, J. I. Basham, T. N. Jackson, O. D. Jurchescu, and D. J. Gundlach, "Mobility overestimation due to gated contacts in organic field-effect transistors," *Nature Commun*. vol. 7, pp. 10908-10914, 2006.
[8] J. Kalinowski and J. Godlewski, "Magnetic-field effects on recombination radiation in tetracene crystal," *Chem. Phys. Lett*. vol. 36, pp. 345-348, 1975.
[9] J. J. Burdett and C. J. Bardeen, "Quantum beats in crystalline tetracene delayed fluorescence due to triplet pair coherences produced by direct singlet fission," *J. Am. Chem. Soc*. vol. 134, pp. 8597-8607, 2012.

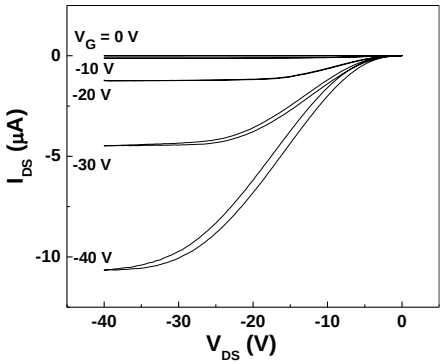


Fig. 1. Measured drain-source current (I_{DS}) *versus* drain-source voltage (V_{DS}) with different gate voltages (V_G) of single crystalline tetracene field-effect transistor under the illumination intensity of 0.34 kW/m^2 at room temperature. Gap between two electrodes was 10 μm.

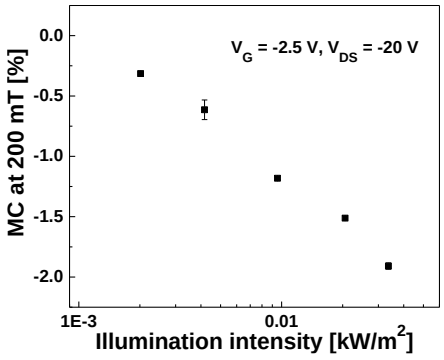


Fig. 2. Measured magnetoconductance (MC) at an external magnetic field of 200 mT *versus* illumination intensity of single crystalline tetracene field-effect transistor at room temperature. Gap between two electrodes was 10 μm.

Jang, Hyuk-Jae; Bittle, Emily; Gundlach, David; Richter, Curt; Zhang, Qin.
"Photo-Induced Magnetic Field Effects in Single Crystalline Tetracene Field-Effect Transistors."
Paper presented at International Semiconductor Device Research Symposium (ISDRS 2016), Bethesda, MD, United States. December 7, 2016 - December 9, 2016.

IMEKO 23rd TC3, 13th TC5 and 4th TC22 International Conference
30 May to 1 June, 2017, Helsinki, Finland

THE NIST MAGNETIC SUSPENSION MASS COMPARATOR: A LOOK AT TYPE B UNCERTAINTY.

C. Stambaugh

National Institute of Standards and Technology, Gaithersburg, MD U.S.A, corey.stambaugh@nist.gov

Abstract: The magnetic suspension mass comparator is a unique system for calibrating kilogram artifacts between vacuum and air. While the magnetic suspension mass comparator allows for direct vacuum-to-air mass measurements, there are several corrections that need to be taken into account. Here, we discuss in greater detail our work to understand the systematic error that results from magnetic interactions with the outside world.

Keywords: Suggest 4-5 keywords.

1. INTRODUCTION

The redefinition of the kilogram in 2018 will alter long established methods for mass dissemination [1]. While the current definition and its subsequent dissemination occurs in air, the new realization will happen in vacuum ($< 10^{-3}$ Pa). This shift is necessary so experiments like the Kibble (Watt) balance [2] and the XRCD project [3] can realize mass at the required levels of precision. For example, the Watt balance relies on interferometry to make precision velocity measurements of the motion of the weighing pan. The variations of the index of refraction in air will affect the measurements and lead to larger uncertainties. However, as one moves down the mass dissemination chain, most end-users will still work at standard atmospheric conditions. The issue lies here, when moving a mass artifact from vacuum to air or air to vacuum the mass value is known to change [4]. The change can be attributed to the adsorption and desorption of molecules in the air from the surface of the artifact. As a result, dissemination methods need to be established for transferring the realized kilogram from vacuum to air to account for these changes. One way of doing this is an established, though indirect method, referred to here as the sorption method. Another way utilizes magnetic suspension, and is aimed towards providing a direct and alternative approach. This paper is focused on the second method, which will allow direct mass comparison between an artifact in vacuum and one in air, and also provide a cross-check on the sorption approach.

The sorption method relies on the repeated measurement of two artifacts of the same material having similar volume but different surface area [5]. The two artifacts are transferred from vacuum to air and back several times. During this period the relative mass change between the two artifacts is measured and an empirical value for the adsorption coefficient can be determined. This value can then be used to account for the mass change per unit surface area when moving an artifact of the same material properties from vacuum to air. Unfortunately, the measured coefficient is known to vary between material types, environment conditions, and the surface quality of the mass artifacts themselves. Thus, repeated measurements and checks for all mass artifacts should be carried out.

The second method, which will be referred to as the magnetic suspension mass comparator or MSMC method provides an alternative by allowing for direct comparisons where material type, quality, etc. are not a factor [6, 7]. The MSMC, see schematic in Fig. 1, is comprised of two, vertically juxtaposed, aluminum chambers. The upper one is typically held under vacuum and will be referred to as the vacuum chamber, while the lower one is held at standard atmospheric conditions and will be referred to as the air chamber. The vacuum chamber houses a 10 kg load commercial mass comparator which has a resolution of 10 μg and a 10 g weighing range. The chamber also houses an apparatus for loading and unloading of masses, both into and out of the chamber and on to and off of the weighing pan of the mass comparator. A special adaptor is connected to the dial weight stack to support the magnetic suspension components. This adaptor holds a pair of downrods that connect to a bridge that is positioned beneath the mass comparator. Hanging from the bridge is a samarium-cobalt (SmCo) magnet surrounded by a tightly wound coil. The coil acts as an electromagnet and provides the variable magnetic force needed for stable suspension of the weighing pan located in the air chamber. Surrounding the coil are a set of magnetic shields used to attenuate the spatial field profile of the SmCo magnet. Directly below the SmCo, but in the air chamber, is an identical magnet with similar magnetic shielding; this is connected to the weighing pan for holding the mass artifacts in air. The air chamber also houses a mass exchange system and holder for positioning the lower magnet assembly and weighing pan in the proper position to achieve suspension. The suspension is covered in more detail by Stambaugh [8].

Briefly, a hall sensor is placed on the flange that separates the vacuum and air chamber and is located directly between the two SmCo magnets. By monitoring the field between the two magnets and feeding back on the electromagnetic coil, suspension of the assembly in the air chamber can be achieved. While magnetically suspended, the entire mass of the sustained suspended assembly is read by the mass comparator in the vacuum chamber. Because the magnetic assembly and weighing pan are suspended during measurements of the mass in vacuum and mass in air, they are ultimately subtracted out of the mass comparison. Thus, one is left with a direct comparison between the masses. Of

Stambaugh, Corey.
"THE NIST MAGNETIC SUSPENSION MASS COMPARATOR: A LOOK AT TYPE B UNCERTAINTY."
Paper presented at IMEKO 23rd TC3, 13th TC5 and 4th TC22 International Conference, Helsinki, Finland. May 30, 2017 - June 1, 2017.

course, there are corrections that must be accounted for and they are discussed below.

Before delving deeper into the MSMC setup and its associated corrections, it is necessary to point out that a similar measurement device exists that is geared toward the measurement of thermodynamic equations of states [9]. Densitometers utilize known weights submerged in a wide range of fluids to measure $p - \rho - T$. That is the density of the fluid, ρ, at different pressures and temperatures. Here, the magnetic suspension is needed because fluids under test can often be at extremely high pressures (>20 MPa) and temperatures (> 400 K), such environments are incompatible with precision mass comparators. The suspension allows for the measurement of the buoyant force acting on the sinker. From this, the density of the fluid under test can be derived. While the magnetic suspension is utilized in a similar manner, that is to measure mass using a mass comparator located in a different chamber, there are several technical differences between the two techniques, which present unique challenges to each. The first of these is the mass comparator range. The densitometer experiments often use a balance with full measurement range and a capacity of 111 g. This allows for more flexibility in the use of calibration weights. Second, the level of precision sought by the density community is at least a factor of 10 less than our desired levels. Finally, the comparison being made there is ultimately between the added fluid in the lower container to a reference fluid. In other words, the comparison is between measurements in the same chamber; we are comparing between two separate chambers.

In order to make an accurate measurement of the mass difference between the mass in vacuum and that in air, three corrections need to be taken into account. The first is the buoyancy correction, as the mass located in air will have an upward buoyant force acting on it that the vacuum mass does not. Second is gravity, as the two masses being compared are located at different heights. The acceleration of gravity varies enough over the distance as to have a measurable effect on the gravitational force acting on the two masses. Finally, there is the force transmission error correction, which relates to how efficiently the gravitational force acting on the suspended assembly is coupled to the balance in the vacuum chamber [9, 10]. Ideally it would be 100 % efficient, but magnetic interactions with the suspended assembly and the outside world are unavoidable. Such interactions must be either minimized or measured so that they can be accounted for. In this paper, we will briefly review the first two corrections and then discuss the origin of the interactions that make the third type of correction necessary, how it impacts us and what we are going to do about it.

2. CORRECTIONS

1. Gravity

The acceleration due to gravity at a point on earth is dependent on the distance from the center of the earth and the density of materials located near the point of interest. To properly account for the change between the two measurement points, we employed a gravimeter to determine the local gravitational gradient. At the location of our experiment, a gradient of $(-2.74 \times 10^{-6} \pm 0.03)\ s^{-2}$ was measured [11]. This will lead to a mass difference of 0.296 mg between the vacuum mass and air mass. The uncertainty of the mass difference is 0.003 mg.

2. Buoyancy

Typically, mass comparisons are made between artifacts under the same environment conditions and corrections are needed to account for differences in volume. In this experiment we are comparing artifacts in two different environments, therefore the buoyancy correction must be taken into account by measuring the volume of the artifact and the air density. At the time of this writing, we have not carried out experiments to verify the expected uncertainty, however we have (a) carried out experiments on mass comparisons between masses of several different volumes and (b) determined the expected uncertainty for our measurement of the density of air. The density of air is determined by measuring pressure, humidity, and temperature while mass comparison measurements are taken. The values are then inserted into standard equations for extracting the air density. We estimate [11, 12] the impact will be 0.013 mg; though improvement is possible.

3. Magnetic Force

Background: The basis for the force transmission error is the magnetic interaction with the outside world; consider a magnet hanging from a balance. The mass of the magnet leads to a gravitational force, F_{g_n}, acting on the balance. The read-out force F_W is

$$F_W = F_{g_n} + F_m(z). \quad (1)$$

The magnet may interact with the outside world and the paramagnetic or diamagnetic interaction force, $F_m(z)$, will respectively add to or subtract from the gravitational force [13], leading to a biased reading. In the example just provided, the systematic error will not affect typical mass calibrations because the difference between the readings of two masses are used. The systematic error gets subtracted out because it does not change; the distance between magnetic and outside world is fixed.

For the case of using magnetic suspension to mediate the connection between the vacuum and air chambers, the interaction problem is slightly more involved, see Fig 1. For this simple derivation, forces resulting from buoyancy and changes in gravity are ignored. The overall force acting on the balance when a mass is placed on the weighing pan directly connected to the balance is

$$F_W = F_{g_n,U} + F_m(z_U) + F_{g_n,L} - F_m\left(z_{L,m_U}\right) + F_{g_n,m_U} \quad (2)$$

and for the mass on the suspended pan

$$F_W = F_{g_n,U} + F_m(z_U) + F_{g_n,L} - F_m\left(z_{L,m_L}\right) + F_{g_n,m_L}. \quad (3)$$

Here $F_{g_n,x}$ is the gravitational force acting on x, where $x = U$ for the assembly located in the vacuum (upper) chamber or x= L for the assembly located in the air (lower) chamber, and m_i is the mass of the artifact in either the upper ($i = U$) or

lower ($i = L$) chambers. $F_m(z_{L,m_i})$ is the magnetic force between the suspended lower assembly and outside world, with ($i = L$) and without ($i = U$) the mass loaded in the lower chamber. Since we are interested in mass differences, we subtract Eq. (2) from Eq. (3):

$$\Delta F_W = \left(F_{g_n,m_U} - F_{g_n,m_L}\right) - F_m(z_{L,m_U}) + F_m(z_{L.m_L}) \quad (4)$$

or

$$\frac{\Delta F}{g} = \Delta m = \Delta m_W + \frac{F_m(z_{L,m_U}) - F_m(z_{L.m_L})}{g}. \quad (5)$$

In this simple model the desired mass value will be shifted by $\delta m = (F_m(z_{L,m_U}) - F_m(z_{L.m_L}))/g$. It is this value that must be minimized. The term δm can be related to the magnetic coupling factor ϕ used in McLinden [10] by the relation, $\phi = 1 + \delta m/m_{m_L}$; m_{m_L} is the mass of the artifact in the lower chamber.

The challenge of accounting for the force transmission error has been covered in depth [9, 10, 14]. However, while instructive, developed methods are not directly applicable to the MSMC and the dissemination of mass. The approach taken for densimeters is to either (a) keep the vertical position of the suspended magnet fixed or (b) measure the effect in situ. To keep the vertical position, $z_{L,m_U} = z_{L,m_L}$, fixed for different mass values, the current can be adjusted to compensate for the change in mass. This poses several problems for the MSMC, as the current, I, needed to compensate for the kilogram change is approximately 1 A: (1) the coil is located in the vacuum so heating is an issue; and (2) the electrical connection to the coil involves high gauge wires that cannot sustain such large current loads. It may be possible to operate at $+I/2$ for one mass value and $-I/2$ for the other. In the steady-state, the average current and thus heating would at least be constant. However, removing the generated heat load from vacuum would still present a challenge. Operating the coil in air, and the suspended pan in vacuum is plausible, but in the current design not straightforward. Then there is still the issue of pushing such a large, constant current through the high gauge wires connecting the coil to the outside word.

Alternatively, in McLinden [10], a prescription is laid out for measuring the effect *in situ*. This approach is not feasible in the MSMC. In short, to correct for the systematic error *in situ* one would need to know the absolute mass values in vacuum and air, which is the very point of the current experiment. If we could start with such knowledge, the system would cease to have utility.

Another approach would be to minimize dF_m/dz. First, this can be achieved by positioning the flange that separates the two magnets closer to the upper magnet. Of course, this has its limits. The separation between the two magnets is fixed, so the distance by which the flange can be moved away from the suspended magnet is finite. Furthermore, if the flange is too close to the coil, variations in the average coil current can also cause interactions. Yet another approach is to minimize the strength of the interaction by decreasing the magnetic permeability of the material between the two magnets.

Finally, we note that another possible source of magnetic interaction comes from the surrounding fluid [10, 14]. The suspended magnetic assembly is kept in a sealed chamber at atmospheric conditions. Approximately 20 % of the air is composed of oxygen. Several groups in the density community have reported force transmission errors resulting from interactions of the suspended magnet with the paramagnetic oxygen in the air.

Results: While corrections for buoyancy and gravity are expected and unavoidable, systematic errors resulting from magnetic interactions are more difficult to predict and correct, and are thus best avoided. Choices such as constructing the chambers using aluminium were done to minimize such effects. When the system was deployed for the very first vacuum to air mass comparison, a systematic error of 92 mg was measured when comparing a mass in the vacuum chamber to one in the air chamber (vacuum-air). When both chambers were held at air (air-air) the effect was 100 mg. After further investigation, it was determined that the difference between the air-air and vacuum-air resulted from the flexing of the aluminium flange separating the two chambers after evacuating the upper chamber. Running suspension without the flange was impractical, so an auxiliary setup was built, where measurements could be made with no flange between the two magnets. In this case, the error dropped to ≈1 mg, which was the resolution of the balance used for the testing. Inserting a thinner aluminium plate between the magnets showed a return of the systematic error, though with a smaller magnitude. Furthermore, the relative position of the flange between the magnets was found to affect the systematic error; the systematic error decreased as the flange was positioned closer to the upper magnet (further from the bottom). All these results are consistent with the expected result indicated by Eq. 5.

In order to solve this problem, there are three options: (1) keep the separation distance between magnets constant, (2) minimize the overall effect, or (3) reduce the magnitude of the change in the systematic error when going from air-air to vacuum-air to a relatively constant value that can be measured accurately and precisely and then corrected for. While the first option is possible, it presents several technical problems. For now, we have decided to explore the latter two options. At the time of this writing, we are testing options and investigating different materials to minimize the magnetic permeability and interaction. By reducing the overall effect, we expect the change from flexing in the flange to decrease significantly, allowing us to follow-up with option (3) above. Air-air measurements can be done to quantify any remaining effect, which could be corrected if sufficiently small. Since the air-air measurements are only concerned with mass differences, the weights used would not require calibration as only their mass values relative to one another would be needed.

Finally, in the measurements we have carried out thus far, we have focused on difference measurements, i.e., we were not configured to accurately measure shifts in mass values, only the differences. To test for any influence from the

oxygen in the environment, we carried out mass comparisons using both room air in the lower chamber and a nitrogen enriched atmosphere in which there was of < 1 % O_2 in the chamber. The upper chamber contained room air at atmospheric pressure. After accounting for the buoyancy difference between humid air and nitrogen, we were unable to measure a difference above the 1 mg level. A further, more detailed measurement and analysis is planned to look for smaller effects.

4. CONCLUSION

The MSMC, through magnetic suspension, allows for a direct comparison between two masses, one in air and the other in vacuum. Like the magnetic suspension systems used for measuring thermodynamic properties of fluids, the magnetic suspension allows the balance to reside in a distinct environment from the measured mass. However, because we are interested in calibrating the mass in air using the mass in vacuum, many of the measurement approaches used by the density community are not applicable; especially when dealing with magnetic interactions between the interface and lower magnet. We have identified the aluminium flange currently separating our two magnets and vacuum from air, as the source of the measured systematic error. Different approaches for dealing with such source of systematic error are currently being investigated, and progress toward this goal will be presented in the talk.

5. REFERENCES

1. Richard, P. and J. Ullrich. *Joint CCM and CCU roadmap towards the redefinition of the SI in 2018*. 2014 [cited 2016 12/12].
2. Stock, M., *Watt balance experiments for the determination of the Planck constant and the redefinition of the kilogram.* Metrologia, 2013. **50**(1): p. R1.
3. Benck, E.C., E. Mulhern, and C. Stambaugh, *Transport of masses under vacuum for the redefinition of the Kilogram at NIST.* Measure, 2016, to be published.
4. Marti, K., P. Fuchs, and S. Russi, *Traceability of mass II: a study of procedures and materials.* Metrologia, 2015. **52**(1): p. 89.
5. Fuchs, P., K. Marti, and S. Russi, *Traceability of mass in air to mass in vacuum: results on the correlation between the change in mass and the surface chemical state.* Metrologia, 2014. **51**(5): p. 376.
6. Abbott, P., et al., *The NIST Mise en Pratique for the Realization and Dissemination of the kilogram as part of the "New SI".* Measure, 2016, to be published.
7. Stambaugh, C. and E. Mulhern, *An FEM analysis of the magnetic fields in the magnetic suspension mass comparator at NIST.* Meausure, 2016, to be published.
8. Stambaugh, C., *The Control System for the Magnetic Suspension Comparator System for Vacuum-To-Air Mass Dissemination.* ACTA IMEKO, 2016, to be published.
9. Wagner, W. and R. Kleinrahm, *Densimeters for very accurate density measurements of fluids over large ranges of temperature, pressure, and density.* Metrologia, 2004. **41**(2): p. S24.
10. McLinden, M.O., R. Kleinrahm, and W. Wagner, *Force Transmission Errors in Magnetic Suspension Densimeters.* International Journal of Thermophysics, 2007. **28**(2): p. 429-448.
11. Mulhern, E. and C. Stambaugh, *Characterization of the NIST Magnetic Suspension Mass Comparator Apparatus and Facility* Measure, 2016, to be published.
12. Picard, A., et al., *Revised formula for the density of moist air (CIPM-2007).* Metrologia, 2008. **45**(2): p. 149.
13. Lösch-Will, C., *Ph. D. Thesis*, in *Lehrstuhls für Thermodynamik*. 2005, Ruhr-Universität: Bochum, Germany.
14. Yuya, K., et al., *A new method for correcting a force transmission error due to magnetic effects in a magnetic levitation densimeter.* Measurement Science and Technology, 2007. **18**(3): p. 659.

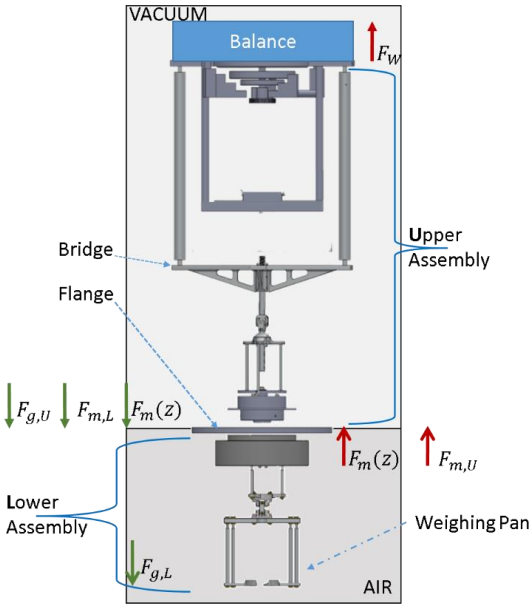


Figure 1: Force diagram for MSMC.

RHEOLOGY OF BALLISTIC CLAY: THE EFFECT OF TEMPERATURE AND SHEAR HISTORY

Ran Tao, Kirk D. Rice, Aaron M. Forster
National Institute of Standards and Technology, Gaithersburg, MD

Abstract

Ballistic clay is used as a backing material for standards-based ballistic resistance tests for the purposes of providing a measure of the energy transferred to the body when a threat is defeated. However, this material exhibits complex thermomechanical behavior under actual usage conditions. In this work, we characterize rheological properties of the standard backing clay material, Roma Plastilina No. 1, used for body armor testing, using a rubber process analyzer. Test methods employed include oscillatory strain sweep, frequency sweep, and oscillatory strain ramp. The results show that the material is highly nonlinear, thermorheologically complex, and thixotropic. The modulus decreases under dynamic deformation and partially recovers when the deformation is discontinued. Experimental protocols developed in this study can be applied for the characterization of other synthetic clay systems.

Introduction

An oil-based modeling clay, Roma Plastilina No. 1[1](RP1), was chosen by both the National Institute of Justice (NIJ) and the Department of Defense (DoD) as the standard backing material for body armor testing since the 1970s [1-3]. Body armor is tested for both penetration and deformation effects (the backface signature) by placing the armor against a backing material, also known as a ballistic witness material (BWM). Therefore, the mechanical and rheological properties of the backing material play a critical role in certification and testing of body armor because most standards allow a maximum indentation depth of 44 mm behind the armor after testing. Unfortunately, over the decades since RP1 was adopted as the standard, changes have been made to the RP1 formulation by the clay manufacturer. Newer versions of RP1 are stiffer at room temperature than the original RP1 selected by the NIJ and the DoD. Ballistics practitioners and researchers must now thermally treat the material prior to use in order to meet clay validation specifications originally developed in the 1970s. Therefore, the ballistic testing community has begun to look for an alternative ballistic witness material to replace RP1 [4,5].

Along with the additional thermal treatment step that makes the verification process of RP1 for ballistic evaluation more cumbersome, another important reason that requires an alternative backing material is the complex rheological behavior of RP1. Specifically, because of its multiphase formulation, the rheological properties of clay are known to be nonlinear in nature, and to depend on work and thermal history, temperature, and time [6-8]. For example, an intuitive observation is that the clay becomes softer after it is worked by hand and becomes harder over time during storage or while it is not used. Such time- and shear-history-dependent complexity is termed as thixotropy. Characterization of a thixotropic material is nontrivial because of the influence of different factors, i.e., time, temperature, and shear history, are coupled in the resulting rheological response. In spite of this, very limited rheological studies have been conducted on RP1 and other clay-like materials. Therefore, the challenges of using BWM are to 1) develop a comprehensive understanding of the rheological properties of the backing material and 2) identify characteristic responses which can be related to backface deformation during the armor ballistic testing.

In this work, we perform rheological studies on the RP1 ballistic clay using a rubber process analyzer (RPA). The RPA is essentially a strain-controlled rotational shear rheometer for rubber testing. The first advantage of the RPA is that is offers a higher torque range for solid materials like clay than that of a commercial rheometer, so that information under larger deformation can be assessed. Second, it provides a consistent sample loading procedure, as ensured by the pneumatic pressure system together with the automatic gap closure, such that the loading effects from sample to sample are minimized.

Materials

RP1 clay was used as received from the manufacturer. The clay bricks from the manufacturer were cut into 70 mm-thick square blocks using a stiff blade. Each sample weighs approximately 5 g to ensure consistency. The sample was placed between two sheets of polyester films and loaded onto the lower die of the RPA for measurements.

1 The full description of the procedures used in this paper requires the identification of certain commercial products and their suppliers. The inclusion of such information should in no way be construed as indicating that such products or suppliers are endorsed by NIST or are recommended by NIST or that they are necessarily the best materials, instruments, software or suppliers for the purposes described.

Tao, Ran; Rice, Kirk; Forster, Aaron.
"Rheology of ballistic clay: the effect of temperature and shear history."
Paper presented at Society of Plastics Engineers Annual Technical Conference (ANTEC), Anaheim, CA, United States. May 8, 2017 - May 10, 2017.

Experimental

Rheological experiments were performed using a rubber process analyzer (RPA). The RPA is equipped with radial serrated bi-cone shape platens with a fixed gap of 0.48 mm. Dynamic strain sweep experiments were performed at 1 Hz from 0.005° strain (equivalent to 0.07 % strain) to 50° strain (equivalent to 697 % strain) at four temperatures of 25 °C, 30 °C, 40 °C, and 50 °C. A fresh sample was used for each test and was loaded at the test temperature. The sample was held for 30 min to allow thermal equilibrium before each run.

Frequency sweep experiments were performed at 0.01° strain (equivalent to 0.14 % strain) and 0.5° strain (equivalent to 7 % strain) from 0.05 Hz to 50 Hz at four temperatures of 25 °C, 30 °C, 40 °C, and 50 °C. To ensure that no additional loading effects or thermal history influence the measurement, the same loading procedure was applied, which is using a fresh sample for each run, loading the sample at the test temperature, and allowing the sample for thermally equilibrate for 30 min prior to measurement.

The breakdown and recovery experiments were performed at 25 °C and 1 Hz using two shear histories. The first consists of increasing the oscillatory shear strain stepwise from 0.01° (equivalent to 0.14 % strain) to 10° (equivalent to 14 % strain) and then decreasing the strain from 10° to 0.01°. The second experiment is to use two shear strain amplitudes of 0.01° and 1° alternatively.

Results

The dynamic shear storage (G') and loss (G'') moduli as a function of strain amplitude at various temperatures for RP1 are shown on a double logarithmic plot in Figure 1. At small strains, the moduli show seemingly linear behavior as the storage modulus barely exhibits a plateau. An examination of the relative intensity of the third harmonic magnitude I_3/I_1 from Fourier analysis of the oscillatory stress waveform (results not shown) show that the strain corresponding to $I_3/I_1 < 3$ % (linear range) is within 0.3 % for all temperatures, which is consistent with the linear range reported by Seppala et al. [7] using a rheometer. As strain increases, both G' and G'' decrease. The modulus values are lower at higher temperature, and the curves show similar shape at different temperatures.

The complex shear modulus G^*, obtained as $|G^*| = \sqrt{G'^2+G''^2}$, is normalized by the complex modulus magnitude at 0.007 % strain for each temperature and plotted as a function of strain amplitude in Figure 2. The modulus curves at different temperature nominally collapse into a single curve for the temperatures examined, indicating similar strain dependence. This

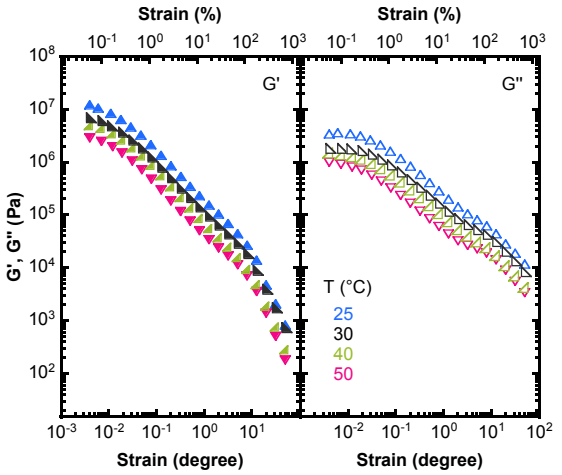


Figure 1. Dynamic shear storage modulus (G', left panel) and loss modulus (G'', right panel) as a function of strain amplitude for RP1 at various temperatures ranging from 25 °C to 50 °C. Standard uncertainty in modulus values associated with this technique is approximately 10 %.

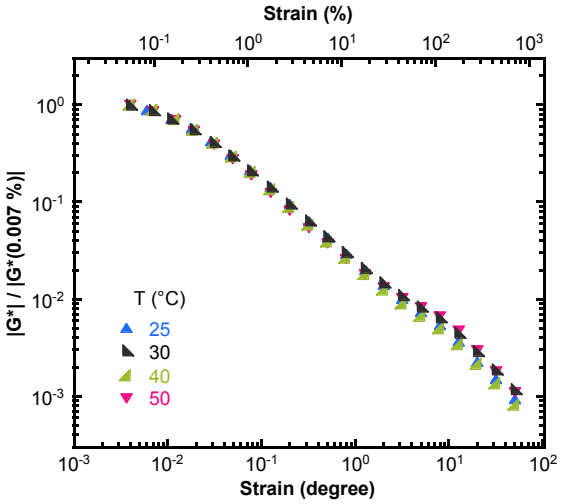


Figure 2. Reduced $|G^*|$ as a function of strain amplitude for RP1 at various temperatures ranging from 25 °C to 50 °C. Standard uncertainty in modulus values associated with this technique is approximately 10 %.

result also suggests that in the measured temperature range, there is no significant thermal transition or structural change that affects the strain dependence of the mechanical behavior.

The results of frequency sweep experiments measured at 0.14 % strain and 7 % strain are shown in Figure 3 and Figure 4, respectively. As expected, the

Tao, Ran; Rice, Kirk; Forster, Aaron.
"Rheology of ballistic clay: the effect of temperature and shear history."
Paper presented at Society of Plastics Engineers Annual Technical Conference (ANTEC), Anaheim, CA, United States. May 8, 2017 - May 10, 2017.

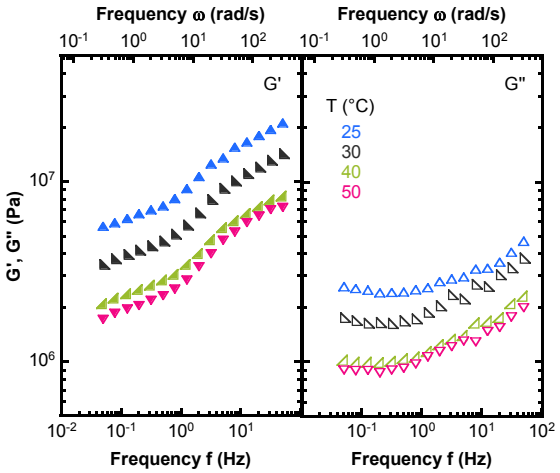


Figure 3. Dynamic shear storage modulus (G', left panel) and loss modulus (G'', right panel) as a function of frequency for RP1 measured at a strain of 0.14 % for various temperatures ranging from 25 °C to 50 °C. Standard uncertainty in modulus values associated with this technique is approximately 10 %.

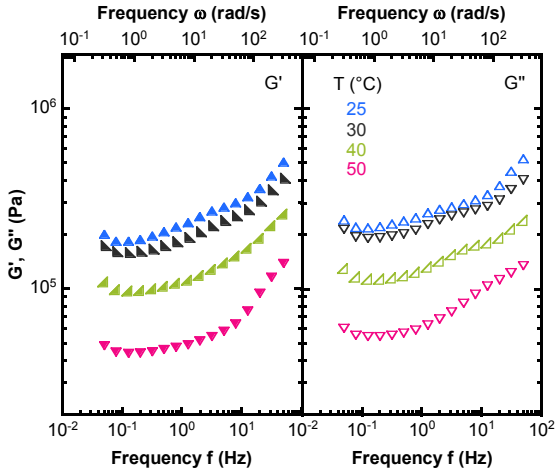


Figure 4. Dynamic shear storage modulus (G', left panel) and loss modulus (G'', right panel) as a function of frequency for RP1 measured at a strain of 7 % for various temperatures ranging from 25 °C to 50 °C. Standard uncertainty in modulus values associated with this technique is approximately 10 %.

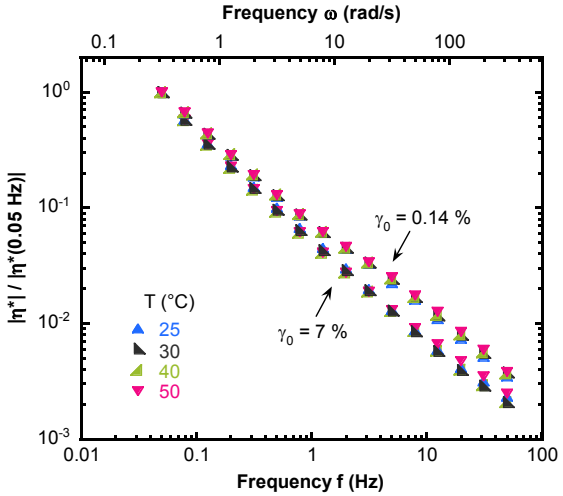


Figure 5. Reduced $|\eta^*|$ as a function of frequency for RP1 measured at strain amplitudes of 0.14 % and 7 % for various temperatures ranging from 25 °C to 50 °C. The arrows show data measured at the corresponding strain. Standard uncertainty in viscosity values associated with this technique is approximately 10 %.

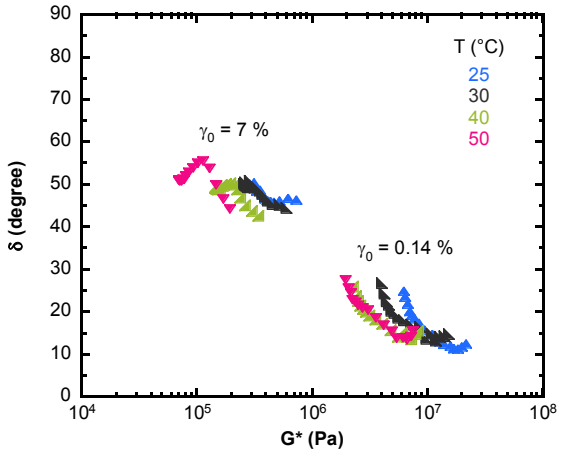


Figure 6. van-Gurp Palmen plot [9] for RP1. The phase angle (δ) in degree is plotted as a function of complex modulus (G^*). Standard uncertainty in modulus values associated with this technique is approximately 10 %.

modulus decreases as temperature increases, and the responses show similar trends at different temperatures. When small deformation is applied (Figure 3), as frequency decreases, G' decreases with a rapid drop first followed by a more moderate decrease, while G'' shows a reduction followed by a plateau. When a large strain of 7 % is applied, the storage and loss moduli share a similar trend, i.e., decrease followed by a plateau, as frequency decreases. To better compare the frequency dependence, the normalized complex viscosity η^*, which is defined as $|\eta^*| = \sqrt{(G'/\omega)^2 + (G''/\omega)^2}$, is plotted in Figure 5 for data measured at those two strain amplitudes. As shown in the figure, the complex viscosity decreases as

Tao, Ran; Rice, Kirk; Forster, Aaron.
"Rheology of ballistic clay: the effect of temperature and shear history."
Paper presented at Society of Plastics Engineers Annual Technical Conference (ANTEC), Anaheim, CA, United States. May 8, 2017 - May 10, 2017.

frequency increases. At a fixed strain, the frequency dependence is the same for all temperatures; however, the complex viscosity shows a greater frequency dependence at higher strain amplitude. This finding suggests that a thorough investigation of the clay behavior requires information over a broad range of both strain amplitude and frequency, because the effects of both factors are involved in the direct impact on the BWM during the verification process of the BWM and ballistic testing.

One way to evaluate if a material is thermorheologically simple or complex is by examining the van Gurp–Palmen plot [9], which is plotting the phase angle (δ) against the corresponding complex shear modulus G^*. For a thermorheologically simple material, within the linear viscoelastic range, the time-temperature superposition (TTS) technique can be used to describe the viscoelastic behavior of the material over a wide range of times or frequencies by shifting the data obtained at different temperatures to a reference temperature [10]. As shown in Figure 6, even for the data obtained at 0.14 %, which is considered to be within the linear region, the curves at different temperature cannot superpose. This result indicates that the RP1 clay is a thermorheologically complex material and TTS does not apply.

Figure 7 shows the results of breakdown and recovery experiments by applying sequential continuous oscillatory shear steps. In the upper figure of Figure 7 where a ramp-up in the shear strain followed by a ramp-down in the shear strain, a breakdown in the complex modulus was observed during the first four steps ($0.01° \rightarrow 0.1° \rightarrow 1° \rightarrow 10°$). When a large strain followed by a small strain is applied ($10° \rightarrow 1°$), the modulus jumps to a higher initial value; however, the modulus continues to decrease during continuous shearing in this step. Moreover, the modulus values for the two pink curves in Figure 7 do not match, indicating that the rheological properties are strongly dependent on previous shear history. Upon further decreasing the strain to 0.1° and to 0.01°, a recovery in the complex modulus is observed. However, the final modulus value dose not reach the initial value, suggesting that some structure in the material has been disrupted and did not recover to the original state within the time of the measurement. Using an alternating oscillatory shear strain of 1° and 0.01°, as shown in the lower figure of Figure 7, an alternating breakdown and recovery in the complex modulus is observed. Strong thixotropy is detected as shown by the evolving modulus during each step. Again, the modulus value is not completely recovered after the first breakdown step.

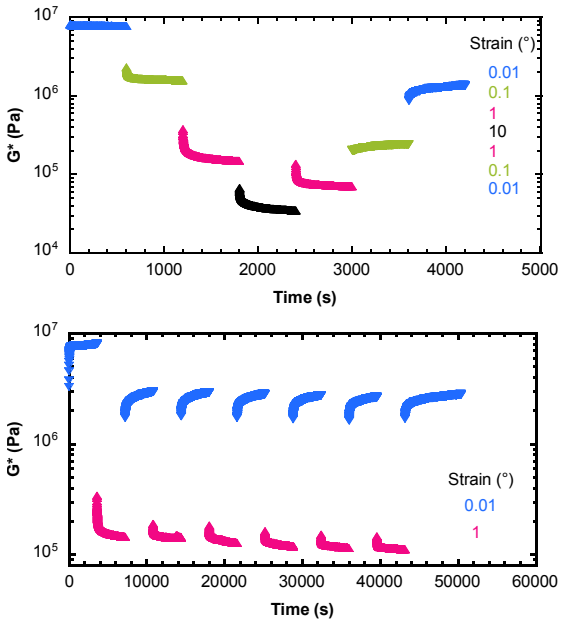


Figure 7. Complex shear modulus G^* as a function of time for the breakdown and recovery experiments. The sequence of applied oscillatory shear strain is from top to down as indicated in the figure. The color of each curve corresponds to the responses obtained from the corresponding strain amplitude is applied. Standard uncertainty in modulus values associated with this technique is approximately 10 %.

Conclusions

In the present work, we studied the effects of temperature and shear history on rheological properties of Roma Plastilina No. 1 clay using a rubber process analyzer. The results show that the clay is a highly nonlinear, thermorheological complex, and thixotropic material. The modulus decreases under dynamic deformation and partially recovers when the deformation is discontinued. Experimental protocols developed in this study are expected to be applied for the characterization of other synthetic clay systems. The continuing clay research at NIST is anticipated to provide fundamental information on the current standard backing material, to extract suitable material parameters that govern the material performance in actual usage conditions, and to help to guide the development of replacement materials.

References

1. R.N. Prather, C.L. Swann, and C.E. Hawkins., Backface Signatures of Soft Body Armors and the

Tao, Ran; Rice, Kirk; Forster, Aaron.
"Rheology of ballistic clay: the effect of temperature and shear history."
Paper presented at Society of Plastics Engineers Annual Technical Conference (ANTEC), Anaheim, CA, United States. May 8, 2017 - May 10, 2017.

Associated Trauma Effects. Aberdeen Proving Ground, Md.: Chemical Systems Laboratory (1977).
2. Ballistic Resistance of Body Armor NIJ Standard-0101.06, US Department of Justice, Office of Justice Programs, National Institute of Justice (2008).
3. DoD Testing Requirements for Body Armor. Report Number D-2009-047, US Department of Defense, Department of Defense Inspector General (2009).
4. National Research Council, Phase II Report on Review of the Testing of Body Armor Materials for Use by the U.S. Army, The National Academies Press: Washington, DC (2010).
5. National Research Council, Phase III Report on Review of the Testing of Body Armor Materials for Use by the U.S. Army. The National Academies Press: Washington, DC (2012).
6. G.B. McKenna, Rheology of Roma Plastilina Clay, National Institute of Standards and Technology: Gaithersburg, MD (1994).
7. J.E. Seppala, Y. Heo, P.E. Stutzman, J.R. Sieber, C.R. Snyder, K.D. Rice, and G.A. Holmes, Characterization of clay composite ballistic witness materials. *J Mater Sci*, **50**, 7048 (2015).
8. D. Bhattacharjee, A. Kumar, I. Biswas, S. Verma and E. Islam, Thermo-Rheological and Dynamic Analysis of Backing Materials for measurement of Behind Armour Blunt Trauma, *Personal Armour Systems Symposium*, Amsterdam (2016).
9. M. van Gurp, J. Palmen, Time-temperature superposition for polymeric blends, *Rheol. Bull.*, **67**, 5 (1998).
10. J.D. Ferry, Viscoelastic Properties of Polymers, 3rd ed.; Wiley: New York (1980).

Tao, Ran; Rice, Kirk; Forster, Aaron.
"Rheology of ballistic clay: the effect of temperature and shear history."
Paper presented at Society of Plastics Engineers Annual Technical Conference (ANTEC), Anaheim, CA, United States. May 8, 2017 - May 10, 2017.

Analytical STEM-in-SEM: Towards Rigorous Quantitative Imaging

Jason Holm[1]

[1.] Applied Chemical and Materials Division, National Institute of Standards and Technology, Boulder, United States.

Modern solid-state transmission electron detectors and conventional scanning electron microscopes are widely available and easy to use, making the suite of techniques referred to as Scanning Transmission Electron Microscopy in a Scanning Electron Microscope (STEM-in-SEM) more accessible today than ever before. These techniques are especially well-suited to a host of applications ranging from nanoparticle metrology [1], to biological or organic sample imaging [2-5], to orientation imaging microscopy [6, 7]. By using low energy primary electrons (i.e., < 30 keV) compared to higher energy primary electrons (i.e. > 80 keV), interaction between the incident electron beam and the sample is more probable, which ultimately means that more information can be obtained. Qualitative image interpretation and extracting quantitative information, however, is not always straightforward since multiple electron scattering events occur in the vast majority of samples. Moreover, rigorous quantitative image analyses can be challenging because electron scattering cross-sections at these energies are not especially well quantified, particularly for low atomic number elements. And, although STEM-in-SEM techniques have existed since the inception of the SEM, very few rigorous experimental transmission imaging studies at conventional SEM energies have been published [8-10].

In this contribution, a straightforward modular aperture system that can be adapted to most STEM detectors will be described [11]. It will be shown how this system can be used to make an angularly sensitive imaging system out of an otherwise rudimentary transmission detector. A rigorous application of that system to thin films comprising materials ranging from carbon to platinum will be demonstrated. From a practical perspective, an improved understanding of image contrast is possible with this system. From a more fundamental perspective, it may be possible to extract electron scattering cross-sections at low energies and for low atomic number elements.

To those ends, systematic experimental studies of several thin films will be presented. Effects of primary electron energy and transmission detector acceptance angle on STEM-in-SEM image intensities will be examined quantitatively and compared with results predicted by Monte Carlo-based electron scattering simulations. Angular distributions of forward scattered electrons will be quantified experimentally using a thin annular detector, ways to obtain imaging conditions that enable rigorous quantitative image interpretation will be described, and mass-thickness contrast at different primary electron energies and detector acceptance angles will be examined.

In one example, commercially available Si_3N_4 films of different thickness (10, 20, and 50 nm) will be examined as a function of primary electron energy and transmission detector acceptance angle. Here, coherent elastic scattering is largely absent and mass-thickness contrast as a function of beam energy is addressed. Angular distributions of electrons forward scattered through the films are quantified and compared with Monte Carlo-based electron scattering simulations, and unanticipated peak splitting in the scattering distributions is observed. In another example, angular distributions of electrons forward-scattered through C, Ni, Co, Pd, and Pt films will be quantified (Figure 1). In addition to quantifying the effect of primary electron energy on the scattering distributions, the following questions will be

Holm, Jason; White, Ryan.
"Analytical STEM-in-SEM: Towards Rigorous Quantitative Imaging."
Paper presented at Microscopy & Microanalysis 2017, St Louis, MO, United States. August 6, 2017 - August 10, 2017.

addressed: Is there an optimal acceptance angle range for quantitative imaging? How easily can elements spanning the periodic table be discerned? Will Pt films always exhibit stronger contrast against vacuum levels than C films if both are the same nominal thickness? Can films of neighboring elements (i.e., Ni and Co) and similar thickness be discerned?

References:
[1] E. Buhr *et al*, Meas. Sci. Technol. **20** (2009) 084025.
[2] M. Kuwajima *et al*, PLOS ONE **8** (2013) e59573.
[3] O. Guise *et al*, Polymer **52** (2011) 1278-1285.
[4] N. Hondow et al, Nanotoxicololgy **5** (2011) 215-227.
[5] C. A. Garcia-Negrete *et al*, Analyst **140** (2015) 3082-3089.
[6] R. Keller and R. Geiss, J. Microscopy **245** (2012) 245-251.
[7] P. Trimby *et al*, Acta Mater. **62** (2014) 69-80.
[8] V. Morandi and P. Merli, J. Appl. Phys. **101** (2007) 114917.
[9] T. Volkenand *et al*, Microsc. Microanal. **16** (2010) 604-613.
[10] M. Pfaff *et al*, J. Microsc. **243** (2011), 31-39.
[11] J. Holm and R. Keller, Ultramicroscopy **167** (2016) 43-56.

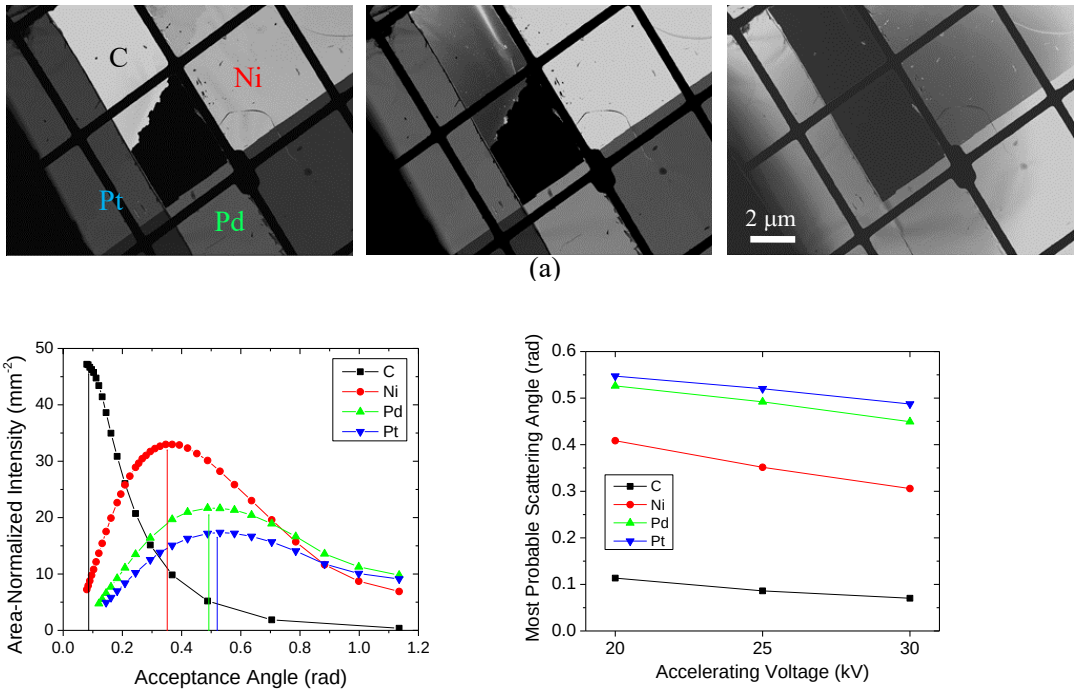


Figure 1. (a) Three STEM-in-SEM images of ~100 nm thick C, Ni, Pd, and Pt films showing contrast changes with increasing detector acceptance angle (acceptance angle increases from left-to-right). (b) Image intensity distributions as a function of detector acceptance angle for 25 keV primary electrons, and (c) A summary of the most probable scattering angles for the four films as a function of primary electron energy.

On Mass-Thickness Contrast in Annular Dark-Field STEM-in-SEM Images

Jason Holm[1] and Ryan White[1]

[1.] Applied Chemical and Materials Division, National Institute of Standards and Technology, Boulder, United States.

A benefit to transmission imaging with low energy (i.e. < 30 keV) primary electrons compared to higher energy (i.e. >100 keV) electrons is that some interactions between the probe electrons and the sample are more likely with decreasing energy. One advantage of the enhanced interaction is that the angular scattering distribution of electrons forward scattered through a sample is likely to broaden with decreasing primary electron energy. By lowering the primary electron energy, samples exhibiting mass-thickness variations may exhibit stronger contrast in some instances, particularly for very thin samples, or for samples comprising low atomic number elements that do not scatter electrons strongly. This is not always the case, however, and depending on the detector acceptance angle employed and the sample composition, images recorded using Scanning Transmission Electron Microscopy in a Scanning Electron Microscope (STEM-in-SEM) techniques can exhibit unanticipated and sometimes confusing contrast [1]. In fact, as will be shown, apparent contrast reversal can be observed even for ultrathin low atomic number samples.

In a recent article, a method to enable comprehensive acceptance angle control for STEM-in-SEM imaging was described [2]. Although comprehensive imaging control is key to extracting the most information from a sample, too much control can also elicit confusing images if the user is unaware of the imaging conditions and electron scattering behavior of their sample. In this contribution, we show that mass-thickness contrast in annular dark-field STEM-in-SEM images can change unexpectedly if the detector acceptance angle is very small and spans a narrow range. The apparent contrast reversal is a function of the detector acceptance angle, and recommendations for setting up imaging conditions for quantitative mass-thickness analyses are offered.

To those ends, an SEM equipped with focused ion beam (FIB) milling capability was used to deposit ultrathin pads of carbon and platinum on ultrathin carbon support films as shown in Figure 1. A 5 keV electron beam was used to deposit discrete pads of increasing thicknesses, and the FIB (Ga^+) functionality was used to mill a small hole adjacent to the pads so that background vacuum image intensity levels could be quantified. Several series of annular dark-field STEM images were recorded of the samples using two small apertures that enabled thin annular detector configurations. The platinum pads were imaged with an annular aperture having inner radius R_i = 0.51 mm and outer radius R_o = 0.78 mm, the carbon pads were imaged using an annular aperture with R_i = 50 μm and R_o = 65 μm. Dark-field images were recorded over a broad range of acceptance angles, and angular distributions of the image intensities were quantified as shown in Figure 2a. In this way, both the acceptance angles at which the maximum image intensities are exhibited and the scattering angle regimes where the apparent contrast reversal occurs are easily discerned. Figure 2b shows a summary of the most probable scattering angles indicated in Figure 2a. A comparison with thickness measurements made using electron energy loss techniques (not shown) concluded that image contrast at the most probable scattering angle is linearly proportional to the sample mass-thickness.

Figure 1 demonstrates that even the simplest of samples can produce images that can be challenging to

Holm, Jason; White, Ryan.
"On Mass-Thickness Contrast in Annular Dark-Field STEM-in-SEM Images."
Paper presented at Microscopy & Microanalysis 2017, St. Louis, MO, United States. August 6, 2017 - August 10, 2017.

interpret. At an acceptance angle of 363 mrad (Fig. 1a), the contrast appears as might be anticipated based on conventional mass-thickness arguments: the thickest pad exhibits the strongest contrast, the thinnest pad exhibits the weakest. At shallower acceptance angles (Fig. 2c) the contrast can be opposite to what might be anticipated: the thinnest pad exhibits the strongest contrast and the thickest exhibits the weakest. At intermediate acceptance angles (Fig. 2b), neither the thickest nor thinnest pads exhibit the strongest contrast. Similar behavior is exhibited by the ultrathin carbon films. Most of the contrast change occurs because of the broadened angular distribution of the forward scattered electrons and the use of small annular apertures that enable only a narrow fraction of the forward scattered electrons to form the images.

The results shown in Figure 2 suggest that when using a narrow annular aperture, the most probable scattering angles (represented by the peaks of the distribution curves) vary directly with pad thickness. Quantitative mass-thickness analyses should be performed at these points under these imaging conditions. Qualitative analyses should be performed at acceptance angles greater than approximately 300 mrad.

References:
[1] V. Morandi and P. Merli, J. Appl. Phys. **101** (2007) 114917.
[2] J. Holm and R. Keller, Ultramicroscopy **167** (2016) 43-56.

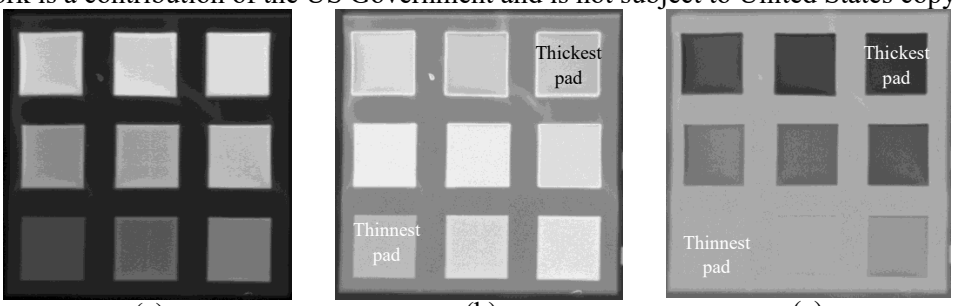


Figure 1. Annular dark-field images of ultrathin platinum pads at (a) $\beta = 363$ mrad, (b) $\beta = 113$ mrad, and (c) $\beta = 42$ mrad. Pads are 2 μm square, and thickness increases from left to right and bottom to top.

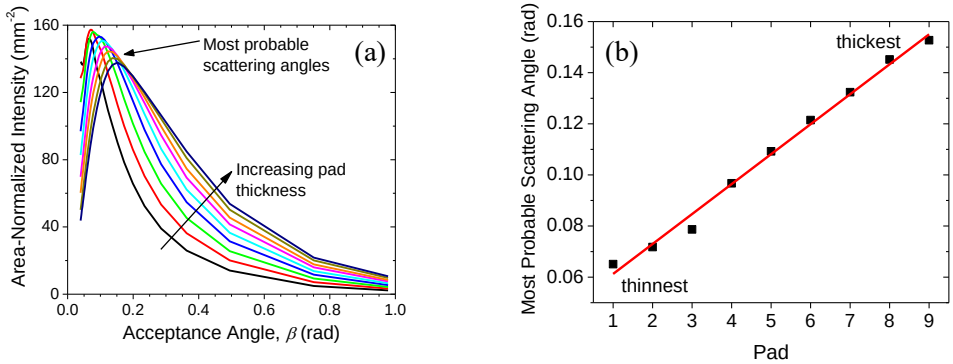


Figure 2. (a) Variation of the image intensity as a function of STEM detector acceptance angle, β, for the platinum sample. (b) The most probable scattering angle as a function of pad thickness. The red line is a guide for the eye.

Characterisation of New Planar Radiometric Detectors using Carbon Nanotube Absorbers under Development at NIST

M.G.White[1], N.A.Tomlin[1], C.Yung[1], M.S.Stephens[1], I.Ryger[1], I. Vayshenker[1], S.I.Woods[2], J.H.Lehman[1]

[1]*The National Institute of Standards and Technology, Boulder, Colorado, USA*
[2]*The National Institute of Standards and Technology, Gaithersburg, Maryland, USA*
Corresponding e-mail address: m.white@boulder.nist.gov

Carbon nanotube technology, in conjunction with silicon micro-fabrication techniques, has enabled us to develop planar radiometric detectors, which has led to the establishment of a new generation of primary standards. The goal is to develop compact, fast, and easy-to-use, radiometric calibration systems, spanning the wavelength spectrum from the ultraviolet to the THz region in a single detector, suitable for use with both coherent and incoherent sources, and encompassing open beam and fibre-coupled modes of operation, with utility beyond that of the laboratory environment. Work will be presented comparing scales derived from two new table-top systems, to existing radiant power and optical fibre power scales, traceable to SI.

INTRODUCTION

NIST has initiated a program titled "*NIST on a Chip"* with the intent of establishing chip-scale standards within key, identified areas of research. The standards themselves are compact and portable, and to an extent ubiquitous. In our case this has been implemented by developing next generation cryogenic and room temperature laser power standards, based on three key enabling technologies.

Vertically aligned CNT arrays with near unity absorptivity, grown on silicon substrates, prepared using state-of-the-art micro-machining techniques, and enhanced with a post growth treatment to stabilise the nanotube array. The core technology has been demonstrated by developing cryogenic radiant power meters [1-3], and room temperature radiometers [4] for both terrestrial and space applications (LASP CSIM). These devices are highly efficient, broad spectrum and two dimensional (planar) in nature. See Figures 1 and 2.

The requirements for a planar detector to be accepted by the community as a viable alternative to a 3 dimensional cavity are twofold. It should demonstrate comparable absorptance, but most importantly long term stability. A cavity by design mitigates reflection losses – a 2 dimensional cavity is not afforded that luxury, unless reflective domes are utilised. It is therefore important to ensure and demonstrate the long term reflectance stability of the VACNT forest. Recent work has concentrated on developing a conformal post-deposition coating process to ensure the deposited nanotubes are characterised as being super-hydrophobic. We have instigated a program to monitor the reflectance change of CNT detectors from the ultraviolet to 10.6 μm.

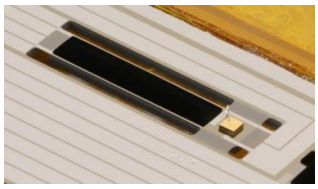


Figure 1. Photograph of a detector for space applications showing the nanotube forest on a silicon membrane.

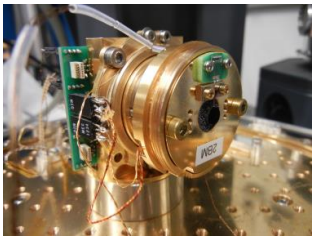


Figure 2. Photograph of a table top mounted cryogenic system showing the black planar detector at the centre. The radiation shields have been removed for clarity. The detector can be used with both open beam and fibre-coupled radiation.

NIST maintains an optical fibre power scale, traceable to SI and disseminated using an electrically calibrated pyroelectric radiometer (ECPR). It is our intention to replace this scale with one derived from our new table-top cryogenic fibre-coupled system,

White, Malcolm; Tomlin, Nathan; Yung, Christopher; Stephens, Michelle; Ryger, Ivan; Woods, Solomon; Lehman, John; Vayshenker, Igor.
"Characterisation of New Planar Radiometric Detectors using Carbon Nanotube Absorbers under Development at NIST."
Paper presented at 13th International Conference on New Developments and Applications in Optical Radiometry (NEWRAD 2017) , Tokyo, Japan. June 13, 2017 - June 16, 2017.

once all characterisation and comparison work is complete. We also maintain spectral power scales on solid-state trap detectors and in the near future we expect to base these scales on our new chip detectors operating in open beam mode in a different, but new, table-top system.

REFERENCES

1. N. A. Tomlin, M. White, I. Vayshenker, S. I. Woods, J. H. Lehman, Planar electrical substitution carbon nanotube cryogenic radiometer, Metrologia **52**, 2, 376-383, (2015).
2. N. Tomlin, A. Curtin, M. White, J. Lehman, Decrease in reflectance of vertically-aligned carbon nanotubes after oxygen plasma treatment, Carbon, **74**, 329-332, (2014).
3. J. Lehman, A. Steiger, N. Tomlin, M. White, M. Kehrt, I. Ryger, M. Stephens, C. Monte, I. Mueller, J. Hollandt, M. Dowell, Planar hyperblack absolute radiometer, Optics Express, **24**, 23, 25911-25921 (2016).
4. I. Ryger, D. Harber, M. Stephens, M. White, N. Tomlin, M. Spidell, J. Lehman, Noise characteristics of thermistors: Measurement methods and results of selected devices, , Rev. of Sci. Instruments, **88**, (2017).

White, Malcolm; Tomlin, Nathan; Yung, Christopher; Stephens, Michelle; Ryger, Ivan; Woods, Solomon; Lehman, John; Vayshenker, Igor.
"Characterisation of New Planar Radiometric Detectors using Carbon Nanotube Absorbers under Development at NIST."
Paper presented at 13th International Conference on New Developments and Applications in Optical Radiometry (NEWRAD 2017) , Tokyo, Japan. June 13, 2017 - June 16, 2017.

Scalable, High-Speed, Digital Single-Flux-Quantum Circuits at NIST

Peter Hopkins, Manuel Castellanos Beltran, Christine Donnelly, Paul Dresselhaus, David Olaya, Adam Sirois, and Sam Benz

Quantum Electrodynamics Division
National Institute of Standards and Technology
Boulder, CO USA

***Abstract*—We have designed, fabricated, and tested niobium-based single-flux quantum (SFQ) digital and mixed-signal circuits based on intrinsically shunted Josephson junctions with tunable niobium-doped amorphous-silicon barriers. This process can be extended to demonstrate dense high-speed SFQ circuits. We have assembled a package of readily-available software design tools for designing, simulating, and optimizing circuits. We have developed a scalable fabrication process at NIST that includes four niobium metal layers and chemical-mechanical planarization of the insulating layers. Through our participation in IARPA's Cryogenic Computing Complexity (C3) program, we have built liquid-helium cryogenic probes and test systems with 40 and 80 input/outputs for characterizing advanced SFQ circuits at speeds up to 26 GHz. Test results of basic SFQ circuits agree with simulations and show large dc bias-current margins.**

Keywords—Rapid single flux quantum (RSFQ); Josephson junctions; Josephson logic; superconducting device fabrication; superconducting integrated circuits

I. Introduction

Superconducting digital logic and interconnects, with compatible cryogenic memories, are being investigated for their potential as an energy-efficient technology for large-scale computing [1]. Recent work has demonstrated significant improvements in the static power dissipation of high-speed single-flux-quantum (SFQ) circuits [2,3] and has quantified the energy efficiency of adiabatic quantum flux parametron (AQFP) circuits [4,5]. Advances in compatible cryogenic memories include switchable Josephson junctions (JJs) consisting of multi-layer barriers containing ferromagnetic materials [6,7], spintronic-based devices scaled to cryogenic temperatures [8-10], and larger arrays of traditional loop-based superconductor memories.

A significant challenge for realizing the potential of these technologies is the ability to geometrically scale the devices so that circuits with millions of JJs and sufficient memory capacity can be fabricated on a standard-sized (~1 cm^2) chip. Both circuit component size requirements and fabrication capabilities are presently limiting these SFQ circuit densities. Components such as shunt resistors for the Josephson junctions and niobium wire inductors contribute significantly to the size of the circuits.

Digital and mixed-signal SFQ circuits with cryogenic memory and based on the propagation of SFQ electrical pulses are of significant interest to NIST for metrology applications related to advanced computing, on-chip signal processing, rf waveform synthesis, and ac voltage standards [11-13]. Significant increases in circuit density will be required to achieve the circuit performance, functionality, complexity, and speed desired for these applications. SFQ circuit components, designs, and materials that are inherently scalable are critical to increasing these circuit densities.

NIST has developed new capabilities to design, simulate, optimize, and fabricate SFQ digital and mixed-signal circuits at the NIST Boulder Microfabrication Facility (BMF). The new fabrication process for these superconducting integrated circuits is an extension of the fabrication process used for our dc and ac voltage standard chips, containing over 265,000 JJs [12,13]. Through our participation on IARPA's Cryogenic Computing Complexity (C3) program, NIST has built cryogenic test infrastructure for measuring and evaluating SFQ circuits of moderate complexity with clock rates up to 26 GHz. This extended abstract describes the NIST SFQ circuit design, fabrication, and test capabilities in more detail.

II. Circuit Design

A. Design Flow

Figure 1 shows the design flow that NIST uses for the SFQ circuits fabricated at the BMF. Commercial electronic design automation (EDA) tools comparable in functionality to those used to design CMOS circuits are not available for designing SFQ circuits [14,15]. Instead, we assembled a package of inexpensive tools that are readily available and have been used by other researchers to successfully design, simulate, and optimize digital SFQ circuits with less than a few thousand Josephson junctions.

 This research is supported by NIST's Innovations in Measurement Science program and is also based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government.

Hopkins, Peter; Castellanos Beltran, Manuel; Donnelly, Christine; Dresselhaus, Paul; Olaya, David; Sirois, Adam; Benz, Samuel.
"Scalable, High-Speed, Digital Single-Flux-Quantum Circuits at NIST."
Paper presented at 2017 16th International Superconductive Electronics Conference (ISEC), Sorrento, Italy. June 12, 2017 - June 16, 2017.

First a circuit schematic is drawn and a pre-layout simulation is performed using tools specifically tailored for superconducting circuits. Circuit components are optimized to maximize the range of dc bias currents (margin) that give correct operation. This is followed by circuit layout and calculation (extraction) of the inductance of the wiring and other circuit components. Design rules specific to our fabrication process, for example the minimum allowable JJ diameter, are checked. The circuit is then simulated and

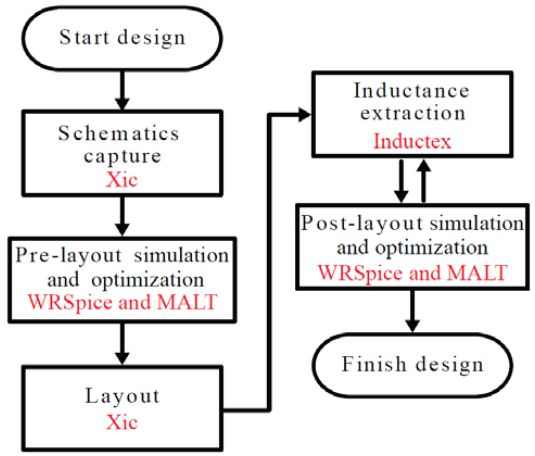


Fig. 1. Design flow for SFQ digital circuits fabricated at NIST.

re-optimized, using the extracted inductance values from the layout. If necessary, the inductance extraction and post-layout simulation cycle is repeated to confirm functionality and maximize current-bias margin. A final design rules check is then done.[1]

B. *Design Details*

Examples of simple rapid SFQ (RSFQ) circuits, such as a Josephson transmission line (JTL), SFQ-to-DC converter, DC-to-SFQ converter, T flip-flop (TFF), and D flip-flop (DFF), were taken from the literature [16]. These designs were modified for our fabrication process and then optimized, using published bias margins as a guide. We then fabricated and tested these circuits and the whole cycle was repeated to further optimize the design and layout for out fabrication process.

As explained below, our fabrication process uses self-shunted junctions, which allows for compact and dense circuits by eliminating the external shunt resistors required for damping junctions and the parasitic inductances associated with shunts. Our JJs are located between layers M1 and M2 (Fig. 2). Designating M3 as the ground plane, closer to the junctions, also reduces parasitic inductances and improves operating margins and operating frequencies [17].

[1]Commercial instruments and design tools are identified in this paper to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the equipment identified are necessarily the best available for the purpose.

III. Fabrication

Circuits are fabricated on oxidized, three-inch silicon wafers. The process consists of four niobium metal layers with chemical-mechanical planarization of the silicon-oxide insulating layers below the third niobium layer. Process cycle time is seven days for one man-shift. More details of the process are given by D. Olaya et al. in a poster at this conference (ISEC 2017).

Our process uses superconductor-normal metal-superconductor (SNS) Josephson junctions with amorphous niobium-doped silicon barriers. These $Nb/Nb_xSi_{1-x}/Nb$ junctions do not require external shunt resistors for damping and proper SFQ circuit operation [18]. The composition and hence the shunt resistance (R_n) and critical current density (J_c) of the barriers can be tuned to selected values of JJ damping, critical current (I_c), area, and characteristic frequency (scaled by the I_cR_n product) as needed for a wide range of applications. The JJs using M1 as the bottom electrode have a targeted J_c of 4.2×10^7 A/m^2, area of ~ 3 μm^2, I_cR_n ~ 0.2 mV, and Stewart-McCumber damping parameter $\beta_c \sim 1$. Future work will target higher J_c, smaller areas, and higher JJ characteristic frequencies to enable higher-speed SFQ circuits.

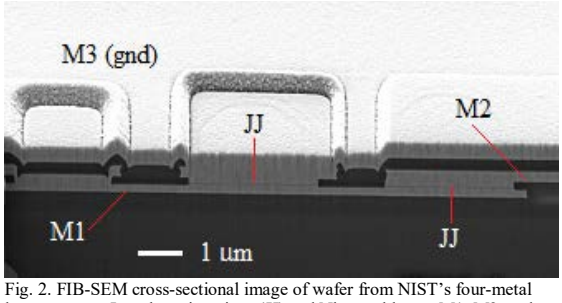


Fig. 2. FIB-SEM cross-sectional image of wafer from NIST's four-metal layer process. Josephson junctions (JJ) and Nb metal layers M1, M2, and M3 are indicated. M0 wiring layer is not shown. Nominal thicknesses are 200 nm, 250 nm, and 350 nm for M1, M2, and M3, respectively.

IV. Test Infrastructure and Results

NIST has built a cryogenic test infrastructure for measuring and evaluating high-speed superconducting digital and mixed-signal circuits. Since our laboratory has a helium recovery and liquefier system, we use immersion probes that are inserted into 100-liter liquid helium storage dewars. Our existing probe has 40 high-speed (26 GHz) input/outputs (I/O) and is capable of handling 5 mm x 5 mm chips. We are also constructing a second probe with 80 high-speed I/Os capable of handling chip or module sizes up to 32 mm x 32 mm.

The 40 I/O probe, as shown in Fig. 3, uses 40 lines of 0.047" semi-rigid coax, with SMA connectors at the room temperature (top) end of the probe and non-magnetic GPPO connectors at the cold end. We have found that preconditioning the coaxial cables by thermally cycling the cables to 200 C prior to installing the connectors is essential for improving the integrity and reliability of the coax-to-GPPO connections. A circular, flexible circuit board printed with forty 50-ohm coplanar waveguides is soldered to the GPPO connectors and

Hopkins, Peter; Castellanos Beltran, Manuel; Donnelly, Christine; Dresselhaus, Paul; Olaya, David; Sirois, Adam; Benz, Samuel.
"Scalable, High-Speed, Digital Single-Flux-Quantum Circuits at NIST."
Paper presented at 2017 16th International Superconductive Electronics Conference (ISEC), Sorrento, Italy. June 12, 2017 - June 16, 2017.

includes 40 pair of beryllium-copper spring fingers that provide press contacts to pads on the perimeter of the chip.

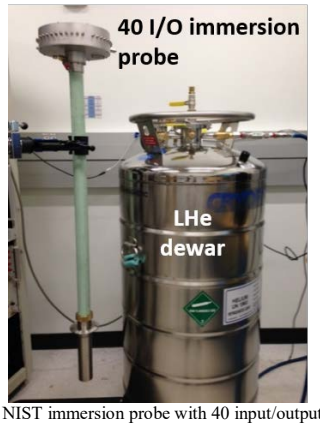


Fig. 3. NIST immersion probe with 40 input/output lines.

We have acquired electronics for low-noise dc and ac biasing of the chips, and high-speed pattern generators and bit error rate detectors for high-speed testing of digital circuits. Circuit functionality, bias margins, bit error rates, and power dissipation can be measured. We have also purchased a commercial electronic test system for biasing and low-speed (1 kHz to 500 kHz) characterization of logic functionality of superconductive circuits. Fig. 4 shows 4 K test results from some of our digital circuits using this low-speed system, comparing the simulated and measured range of dc bias current values for which the circuits work correctly. The horizontal axis is the deviation of the bias current (I) from the design bias value (I_b), normalized by I_b (i.e. $(I-I_b)/I_b$).

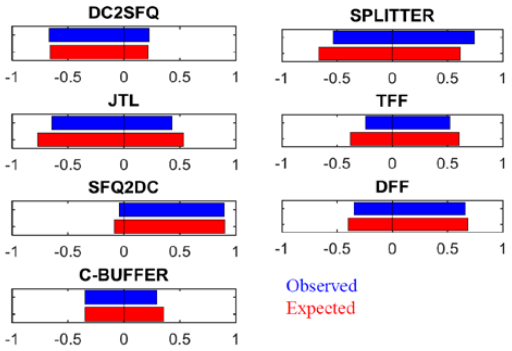


Fig. 4. Simulated vs. 4 K measured dc bias current margins (I-Ib)/Ib for our SFQ digital circuits. Measurements used 1 kHz clock rate.

ACKNOWLEDGMENTS

We acknowledge High Precision Devices for design consultation and construction of our cryogenic test probes.[1] We thank Igor Vernik and Oleg Mukhanov at Hypres, Josh Osborne at Northrop Grumman for helpful discussions on probe design and testing of SFQ circuits, and Stephen Whiteley at Whiteley Research for help with design simulation software.

REFERENCES

[1] D.S. Holmes, A.L. Ripple, and M.A. Manheimer, "Energy-Efficient Superconducting Computing–Power Budgets and Requirements," IEEE Trans. Appl. Supercond., vol. 23, pp. 1701610, June 2013.

[2] A.Y. Herr, Q. P. Herr, O. T. Oberg, O. Naaman, J. X. Przybysz, P. Borodulin, and S. B. Shauck, "An 8-bit carry look-ahead adder with 150 ps latency and sub-microwatt power dissipation at 10 GHz," J. Appl. Phys., vol. 113, no. 3, pp. 033911-1–033911-6, Jan. 2013.

[3] D.E. Kirichenko, S. Sarwana, and A. F. Kirichenko, "Zero static power dissipation biasing of RSFQ circuits," IEEE Trans. Appl. Supercond., vol. 21, no. 3, pp. 776–779, June 2011.

[4] N. Takeuchi, D. Ozawa, Y. Yamanashi and N. Yoshikawa, "Adiabatic quantum flux parametron as an ultra-low-power logic device," Supercond. Sci. Tech., 26, 035010, Jan. 2013.

[5] N. Takeuchi, Y. Yamanashi and N. Yoshikawa, "Measurement of 10 zJ energy dissipation of adiabatic quantum-flux-parametron logic using a superconducting resonator," Appl. Phys. Lett., 102, 052602, Feb. 2013.

[6] A.I. Buzdin, L.N. Bulaevskij, and S.V. Panyukov, "Critical-current oscillations as a function of the exchange field and thickness of the ferromagnetic metal (F) in an S-F-S Josephson junction," J. Exp. Theor. Phys. Lett. 35, 178–180, 1982.

[7] V.V. Ryazanov, V.A. Oboznov, A. Yu. Rusanov, A.V. Veretennikov, A.A. Golubov, and J. Aarts, "Coupling of Two Superconductors through a Ferromagnet: Evidence for a π Junction," Phys. Rev. Lett. 86, 2427-2430, March 2001.

[8] C. Bell, G. Burnell, C.W. Leung, E.J. Tarte, D.-J. Kang, and M. G. Blamire, "Controllable Josephson current through a pseudospin-valve structure," Appl. Phys. Lett. 84, 1153–1155, Feb. 2004.

[9] B. Baek, W.H. Rippard, S.P. Benz, S.E. Russek, and P.D. Dresselhaus, "Hybrid superconducting-magnetic memory device using competing order parameters," Nat. Commun. 5, 3888, May 2014.

[10] E.C. Gingrich, B.M. Niedzielski, J.A. Glick, Y. Wang, D.L. Miller, R. Loloee, W.P. Pratt, and N.O. Birge, "Controllable 0–π Josephson junctions containing a ferromagnetic spin valve," Nat. Phys. 10, 1038, March 2016.

[11] J.A. Brevik et al., "Josephson arbitrary waveform synthesis with multilevel pulse biasing," IEEE Trans. Appl. Supercond., vol. 27, no. 3, 1301707, April 2017.

[12] P.D. Dresselhaus, M.M. Elsbury, D. Olaya, C.J. Burroughs, and S.P. Benz, "10 Volt Programmable Josephson voltage standard circuits using NbSi-barrier junctions," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 693 - 696, June 2011.

[13] A. E. Fox, P. D. Dresselhaus, A. Rüfenacht, A. Sanders, and S. P. Benz, "Junction Yield Analysis for 10 V Programmable Josephson Voltage Standard Devices," IEEE Trans. Appl. Supercond., vol. 25, no. 3, pp. 1101505-5, June 2015.

[14] C.J. Fourie and M.H. Volkmann, "Status of Superconductor Electronic Circuit Design Software," IEEE Trans. Appl. Supercond., vol. 23, no. 3, 1300205, June 2013.

[15] A. Inamdar, J. Ren, and D. Amparo, "Improved Model-to-Hardware Correlation for Superconductor Integrated Circuits," IEEE Trans. Appl. Supercond., vol. 25, no. 3, 1-8, June 2015.

[16] K.K. Likharev and V.K. Semenov, "RSFQ logic/memory family: A new Josephson-junction technology for sub-terahertz-clock-frequency digital systems," IEEE Trans. Appl. Supercond., vol. 1, pp 3-28, March 1991.

[17] M. Maezawa, "Numerical Study of the Effect of Parasitic Inductance on RSFQ Circuits," IEICE Trans. Electr., vol. E84-C, no. 1, pp 20-28, Jan. 2001.

[18] D. Olaya, P.D. Dresselhaus, and S.P. Benz, "300-GHz operation of divider circuits using high-Jc Nb/NbxSi1-x/Nb Josephson junctions," IEEE Trans. Appl. Supercond., vol. 25, no. 3, 1101005, June 2015.

Hopkins, Peter; Castellanos Beltran, Manuel; Donnelly, Christine; Dresselhaus, Paul; Olaya, David; Sirois, Adam; Benz, Samuel.
"Scalable, High-Speed, Digital Single-Flux-Quantum Circuits at NIST."
Paper presented at 2017 16th International Superconductive Electronics Conference (ISEC), Sorrento, Italy. June 12, 2017 - June 16, 2017.

Fabrication of High-Speed and High-Density Single-Flux-Quantum Circuits at NIST

D. Olaya, P. D. Dresselhaus, P. F. Hopkins, S. P. Benz
Quantum Electromagnetics Division
National Institute of Standards and Technology
Boulder, CO USA

***Abstract*—The development of a fabrication process for single-flux-quantum (SFQ) digital circuits is a fundamental part of the NIST effort to develop a gigahertz waveform synthesizer with quantum voltage accuracy. This paper describes the current SFQ fabrication process used at NIST's Boulder Microfabrication Facility based on Josephson junctions with niobium superconducting electrodes and self-shunted niobium-doped silicon barriers. The planned circuits will be used to explore metrology for advanced computing and communications. We will also describe anticipated process changes and innovations aimed toward further increasing circuit density and clock frequency.**

Keywords—Josephson junctions, self-shunted junction, RSFQ, superconducting integrated circuit, superconductor electronics

I. Introduction

For over 20 years, researchers at NIST have been developing quantum-accurate waveform synthesizers for ac voltage metrology. These voltage sources are fundamentally accurate in voltage amplitude and have perfect signal purity (distortion free) [1]. State-of-the-art circuits made with series arrays of Josephson junctions (JJs) have generated waveforms with frequencies up to 1 MHz with rms amplitude up to 2 V **Error! Reference source not found.**. Semiconductor pulse generators have been used to bias the long arrays of overdamped, self-shunted, SNS Josephson junctions. The goal now is to extend quantum-accurate waveform synthesis to gigahertz frequencies to provide advanced calibration tools for the telecommunications industry. One approach to accomplish this is by using digital SFQ circuits to directly generate digitally synthesized waveforms because it allows faster clock rates that will in turn enable synthesis of higher-performance oversampled waveforms. The first SFQ circuits we are developing will synthesize waveforms at rf frequencies at amplitudes of a few millivolts.

New technology is needed to extend quantum-accurate waveform generation into the gigahertz regime. The digital synthesis that is presently done using semiconductor electronics, needs to be clocked faster by at least a factor of 10, preferably at least 100 GHz. This will be achieved by replacing those semiconductor components with faster superconducting SFQ digital electronics **Error! Reference source not found.**. SFQ signal processing and waveform synthesizer circuits will be fabricated on the same chip.

II. Fabrication Process

The circuits are fabricated on three-inch silicon wafers covered with 150 nm of thermal oxide. The fabrication process of these circuits includes four niobium layers with partial planarization on the insulating layers below the third niobium layer, as shown schematically in Fig. 1. Either the bottom or top niobium layers may be utilized as the ground plane, as shown in Fig. 2. Wiring layers are made of niobium sputtered in argon and etched using sulphur hexafluoride reactive ion etching (RIE).

An important and distinctive feature of our fabrication process is the broad tunability of the electrical properties of our junctions. Our barriers, which consist of co-sputtered niobium and silicon, are adjusted by changing the sputter rate of the

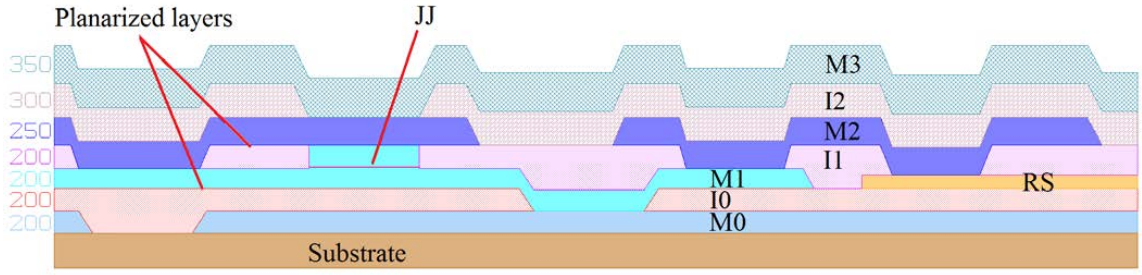


Fig. 1. Schematic of cross section of NIST SFQ digital fabrication process. M# designates niobium layers, I# designates insulator layers, RS is the resistors layer, JJ is the self-shunted junction with niobium silicide barrier.

This work is a contribution of the U.S. government and is not subject to U.S. copyright. This research was supported by the NIST Innovations in Measurement Science program and based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government.

Olaya, David; Dresselhaus, Paul; Hopkins, Peter; Benz, Samuel.
"Fabrication of High-Speed and High-Density Single-Flux-Quantum Circuits at NIST."
Paper presented at 2017 16th International Superconductive Electronics Conference (ISEC), Sorrento, Italy. June 12, 2017 - June 16, 2017.

niobium and the barrier thickness. For waveform synthesizer applications, we tune the barrier so that the junctions are damped and their characteristic frequency allows for fast operation of digital SFQ circuits **Error! Reference source not found.**. Self-shunting of the junctions is critical for scaling future circuits to higher density because it avoids external shunt resistors that take significant chip area. It is also important for achieving faster circuits by eliminating the parasitic inductances associated with shunts. The current process yields junctions with a critical

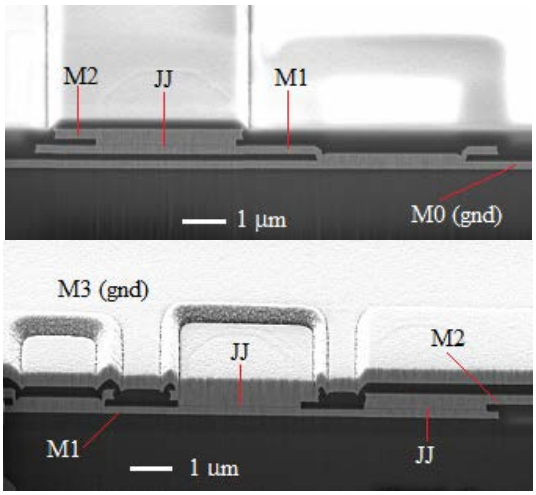


Fig. 2. Cross section images of junctions, top image shows JJ grounded to M0, bottom image shows JJ grounded to M3.

current density (J_c) of 4.2 kA/cm^2. This value of J_c was chosen as the initial step to develop the cell library and the fabrication process.

Insulator layers consist of silicon oxide deposited by plasma enhanced chemical vapor deposition (PECVD). Back side helium flow during deposition ensures that the wafer temperature remains under 100 C. Planarization, which is done on insulating layers I0 and I1, primarily follows the typical caldera method **Error! Reference source not found.**. First, an oxide of 500 nm thickness is deposited to cover the patterned metal. This is covered with photoresist in areas around where the underlying metal has been etched. By timed etch of the oxide (200 nm), calderas are left over a flat surface. Next, chemical-mechanical polishing (CMP) removes 100 nm of the remaining oxide leaving a flat surface of (200 nm)-thick oxide over the metal and (400 nm)-thick oxide over the previous insulator, as shown in Fig. 3. Thickness accuracy after the process is +/- 20 nm.

Resistors made of palladium-gold alloy with sheet resistance of 2 Ω per square are e-beam evaporated on a bi-layer of lift-off resist and imaging resist. The contact pads are made with a second sputter deposition of palladium-gold. These pads

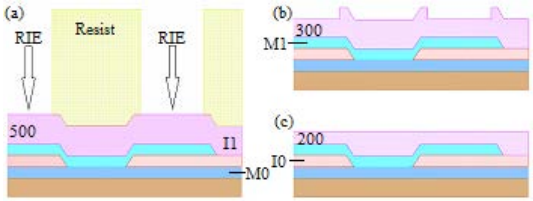


Fig. 3. Steps for planarization of insulator layers. (a) Extra thickness of insulator deposited, resist covering areas lower than surroundings, RIE. (b) After RIE calderas are etched. (c) CMP removes caldera protrusions and an additional 100 nm of oxide, leaving a flat surface.

are lifted off with a single layer resist because the size tolerances for these features are more relaxed. Table I shows the thicknesses and materials of each of the layers in the process.

III. Process Developments

Towards the goal of producing quantized waveforms in the gigahertz frequency range, it will be necessary to make SFQ circuits clocked at hundreds of gigahertz. This requires maximizing the critical current density, J_c, of the junctions, since the operating speed of these circuits scales approximately as $\sqrt{J_c}$ for a fixed JJ damping **Error! Reference source not found.**.

TABLE I. Layers

Layer	Thickness (nm)	Material	Minimum feature size (µm)
M0	200	Niobium (sputtered)	0.5
I0	200	SiOx (PECVD)	0.75
M1	200	Niobium (sputtered)	0.75
JJ	10	Niobium silicide (cosputtered)	1.0
I1	200	SiOx (PECVD)	0.6
RS	135	Palladium-gold (evaporated)	0.75
M2	250	Niobium (sputtered)	1.0
I2	300	SiOx (PECVD)	1.0
M3	350	Niobium (sputtered)	1.0
PADS	200	Palladium-gold (sputtered)	2.0

The process that aims to maximize circuit operating speeds will use junctions with J_c = 100 kA/cm^2 and defined by electron-beam lithography. For this J_c, the JJ sizes need to decrease to deep submicrometer dimensions, which is beyond the capabilities of the stepper used in the current process, to maintain adequate critical current values.

Olaya, David; Dresselhaus, Paul; Hopkins, Peter; Benz, Samuel.
"Fabrication of High-Speed and High-Density Single-Flux-Quantum Circuits at NIST."
Paper presented at 2017 16th International Superconductive Electronics Conference (ISEC), Sorrento, Italy. June 12, 2017 - June 16, 2017.

SFQ circuits with self-shunted Nb-Si junctions and J_c = 85 kA/cm^2 have already been demonstrated at NIST [7]**Error! Reference source not found.**. The maximum operating frequency of those circuits was 300 GHz. We believe that the maximum frequency of those circuits was limited by both the fabrication process and the design tools used at that time, neither of which were optimal for these circuits. For the following reasons, we anticipate better circuit performance with the new process, including better yield and an increase in the maximum operation frequency. New design tools have been implemented that have allowed circuit parameters to be optimized for higher speeds. A new e-beam writer, recently acquired by the Boulder Microfabrication Facility, has enabled smaller, more uniform features sizes that should translate into better control of JJ critical currents and better uniformity. Incorporation of CMP will allow for better contact to smaller junctions and increased current density of writing interconnects.

Future plans to increase circuit density include investigating the replacement of geometric inductors with the Josephson inductance of non-switching vertically stacked junctions [8]. A modified fabrication process would permit a switching junction and a stacked-JJ inductor to reside within the same vertical stack, almost eliminating all the area occupied by a geometric inductor. The combination of small, high-J_c, self-shunted junctions and stacked JJ inductors will enable a significant increase in SFQ circuit density.

Other planned improvements to the process include the introduction of stud vias, which will improve the connection between metal layers by decreasing parasitic inductance, and improving planarization, which will allow the addition of more metal layers as the circuits increase in complexity.

REFERENCES

[1] S. P. Benz and C. A. Hamilton, "A pulse-driven programmable Josephson voltage standard," Appl. Phys. Lett., vol. 68, pp. 3171–3173, May 1996.

[2] J. A. Brevik et al., "Josephson arbitrary waveform synthesis with multilevel pulse biasing," IEEE Trans. Appl. Supercond. vol. 27, no. 3, 1301707, April 2017.

[3] W. Chen, A. V. Rylyakov, V. Patel, J. E. Lukens, and K. K. Likharev, "Rapid single flux quantum T-flip flop operating up to 770 GHz," IEEE Trans. Appl. Supercond., vol. 9, pp. 3212-3215, June 1999.

[4] B. Baek, P. D. Dresselhaus, and S. P. Benz, "Co-sputtered amorphous Nb_xSi_{1-x} barriers for Josephson-junction circuits," IEEE Trans. Appl. Supercond., vol. 16, pp. 1966-1970, Dec. 2006.

[5] S. Nagasawa et al., "Planarized multi-layer fabrication technology for LTS large-scale SFQ circuits," Supercond. Sci. Technol., vol. 16, pp. 1483 – 1486, Nov. 2003.

[6] S. K. Tolpygo et al., "20 kA/cm^2 Process Development for Superconducting Integrated Circuits with 80 GHz Clock Frequency", IEEE Trans. Appl. Supercond., vol. 17, pp. 946-951, June 2007.

[7] D. Olaya, P. D. Dresselhaus, and S. P. Benz., "300-GHz operation of divider circuits using high-J_c Nb/Nb_xSi_{1-x}/Nb Josephson junctions," IEEE Trans. Appl. Supercond., vol 25, no. 3, 1101005, 2015.

[8] Oleg ISEC talk, and private communiction, 2015.

Olaya, David; Dresselhaus, Paul; Hopkins, Peter; Benz, Samuel.
"Fabrication of High-Speed and High-Density Single-Flux-Quantum Circuits at NIST."
Paper presented at 2017 16th International Superconductive Electronics Conference (ISEC), Sorrento, Italy. June 12, 2017 - June 16, 2017.

Analytical description of charged grain boundary recombination in polycrystalline thin film solar cells

Benoit Gaury
Center for Nanoscale Science and Technology
National Institute of Standards and Technology
Gaithersburg, MD, 20899, USA
Maryland NanoCenter, University of Maryland
College Park, MD 20742

Paul M. Haney
Center for Nanoscale Science and Technology
National Institute of Standards and Technology
Gaithersburg, MD, 20899, USA

***Abstract*—We present analytic expressions for the dark current-voltage relation of a pn^+ junction with a positively charged columnar grain boundary, containing a distribution of defect states in the band gap. A closed form relation for the open-circuit voltage $V_{\rm oc}$ is provided for an illuminated junction. These findings are verified by direct comparison with numerical simulations, and provide a quantitative understanding of the reduction of $V_{\rm oc}$ by grain boundaries in thin film photovoltaic materials.**

***Index Terms*—grain boundary, thin films, recombination.**

I. Introduction

Despite years of research, precise guidelines for the improvement of chalcogenide materials, like CdTe, are still unavailable for photovoltaic applications [1]. In particular, the role of grain boundaries in these materials is not well understood. High densities of grain boundaries generally enhance recombination, hence reducing power conversion efficiency. However, recent development in thin-film photovoltaics based, for instance, on CdTe or Cu(In,Ga)Se_2, have led to unexpectedly high efficiencies despite large densities of grain boundaries [2].

Even though nanaoscale measurements may be difficult to interpret [3], electron beam induced current [4]–[6] and Kelvin probe microscopy [7]–[9] experiments have revealed positively charged grain boundaries in these materials. So far, the corresponding body of theoretical studies has consisted of one-dimensional models [10]–[13], and more complex numerical simulations [14]–[17]. The latter revealed the paradoxical impact of grain boundaries on the efficiency of thin film solar cells. While grain boundaries might improve carrier collection, they also reduce the open-circuit voltage, leading to a more contrasted picture than often presented. Our recent analytical works Refs. [18], [19] are consistent with this finding.

In this work, we present new analytical expressions for the dark current-voltage $J(V)$ relation of a pn^+ junction with positively charged grain boundaries, containing many discrete defect states in the gap. We use physical descriptions of the electron and hole transport to create a simplified model which captures the essential features of the drift-diffusion-Poisson equations. The accuracy of this model and the corresponding predictions for the grain boundary dark recombination current are verified numerically.

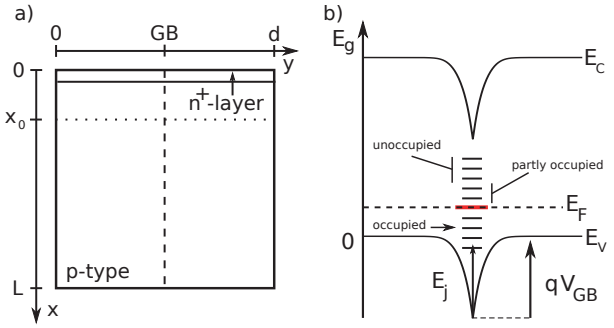


Fig. 1. (a) Two-dimensional model system of a pn^+ junction containing a single grain boundary. (b) Band structure in the neutral region of the p-doped semiconductor, orthogonal to the grain boundary. The dashed line is the thermal equilibrium Fermi level E_F, surrounded by the grain boundary defect states E_j. E_C and E_V are the maxima of the conduction and valence bands, E_g is the material bandgap energy, $V_{\rm GB}$ is the grain boundary built-in potential. The red line indicates the defect neutral energy level $E_{\rm GB}$. We take the energy reference at the valence band edge in the bulk of the neutral region.

II. Physical model of a grain boundary with multiple gap states

Our model system, depicted in Fig. 1(a), is a pn^+ junction with a vertical grain boundary. We assume selective contacts such that the hole (electron) current vanishes at the $n(p)$-contact, and we use periodic boundary conditions in the y-direction. We note x_0 the position where electron and hole densities are equal in the grain interior.

Polycrystalline semiconductors possess a wide variety of defects, which all have specific formation energies as a function of the local environment and Fermi level. For a given Fermi level, we assume that the type of defect with the lowest formation energy is the most prevalent. Our model then considers a distribution of gap states that results from this particular defect. Because experimental evidence tend to indicate that grain boundaries are positively charged in materials like CdTe [4], [7], [8], we focus on distributions of gap states that provide positively charged defects. In order to conserve the electroneutrality of the device, the positive charges must be screened by nearby negative charges: free electrons in an n-type material or ionized acceptor dopants in

Gaury, Benoit; Haney, Paul.
"Analytical description of charged grain boundary recombination in polycrystalline thin film solar cells."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

a p-type material. In a p-type material, the system responds to the positive charge by developing an electric field around the grain boundary. This electric field repels holes from the grain boundary core, creating a depleted region around it. This depleted region is negatively charged because of the uncompensated ionized acceptors, and therefore compensates the positive charge of the defect. For the band structure across the grain boundary shown in Fig. 1(b), the electric field results in the formation of a built-in potential $V_{\rm GB}$ around the grain boundary. The amplitude of $V_{\rm GB}$ depends on the doping density, the distribution and density of the gap states.

The built-in potential and the Fermi level determine the type of the grain boundary, i.e. the majority carrier at the grain boundary core. In a p-type material, low values of built-in potentials create a hole depletion at the grain boundary, but holes remain majority carrier at the grain boundary core. Large $V_{\rm GB}$ values lead to type inversion at the grain boundary core, i.e. electrons become majority carrier. Each grain boundary type has a different carrier transport under nonequilibrium conditions. In this work we consider both inverted and non-inverted grain boundaries.

We model the grain boundary as a two-dimensional plane with amphoteric (acceptor and donor) states. The grain boundary charge reads

$$Q_{\rm GB} = q\rho_{\rm GB}\sum_{j=1}^{M}(1-2f_{\rm GB}(E_j)) \tag{1}$$

where $\rho_{\rm GB}$ is the 2D defect density of states, and M is the number of gap states. The occupancies of each state is given by

$$f_{\rm GB}(E_j) = \frac{S_n n_{\rm GB} + S_p \bar{p}_j}{S_n(n_{\rm GB}+\bar{n}_j) + S_p(p_{\rm GB}+\bar{p}_j)}, \tag{2}$$

where $n_{\rm GB}$ ($p_{\rm GB}$) is the electron (hole) carrier density at the grain boundary, S_n (S_p) is the electron (hole) surface recombination velocity, $\bar{n}_j$ and $\bar{p}_j$ are

$$\bar{n}_j = N_C e^{(-E_g+E_j)/k_BT} \tag{3}$$

$$\bar{p}_j = N_V e^{-E_j/k_BT} \tag{4}$$

where E_j is a defect energy level calculated from the valence band edge, N_C (N_V) is the conduction (valence) band effective density of states, E_g is the material bandgap, k_B is the Boltzmann constant and T is the temperature. At thermal equilibrium Eq. (2) reduces to the Fermi-Dirac distribution $f_{\rm GB}(E) = (1+\exp[(E-E_F)/k_BT])^{-1}$. We introduce $E_{\rm GB}$ as the neutral level of the distribution of states, such that the occupied and empty gap states (below and above $E_{\rm GB}$) contribute an equal and opposite charge (red line in Fig. 1(b)). In thermal equilibrium, the problem of many defect states can therefore be mapped onto an effective *single* defect level positioned at energy $E_{\rm GB}$. The charge of the single effective level is then

$$Q_{\rm GB} = qM\rho_{\rm GB}(1-2f_0(E_{\rm GB})), \tag{5}$$

where f_0 is the effective occupancy of the effective state $E_{\rm GB}$. In the limit of large defect density of states, we find that $E_{\rm GB}$ is given by the average of the gap state energies for an even number of states, and is equal to the gap state that has the same number of states above and under it for an odd number of states.

We consider large grain boundary defect densities such that the Fermi level is pinned at $E_{\rm GB}$ (see Fig. 1(b)). This situation occurs for densities above the critical value

$$\rho_{\rm GB}^{\rm crit} = \frac{1}{q}\frac{\sqrt{8\epsilon N_A(E_{\rm GB}-E_F-k_BT)}}{\sum_{j=1}^{M} 1-\frac{2}{1+e^{(E_j-E_{\rm GB}+k_BT)/k_BT}}}. \tag{6}$$

Defining $V_{\rm GB}^0$ as the equilibrium potential between the grain boundary and bulk of the neutral region, then assuming $\rho_{\rm GB} > \rho_{\rm GB}^{\rm crit}$ leads to

$$qV_{\rm GB}^0 \approx E_{\rm GB} - E_F. \tag{7}$$

The scope of this work is limited to built-in potentials such that $V_{\rm GB}^0 \gg k_BT/q$.

III. Assumptions and grain boundary properties away from equilibrium

The full two-dimensional drift-diffusion-Poisson set of equations is not analytically solvable. We therefore focus on the solution along the grain boundary core, which reduces the problem to one dimension. This is made possible by the electrostatic confinement of electrons to the grain boundary core. Our analysis relies on assumptions that enable analytical solutions.

Our main assumption is that the hole quasi-Fermi level is flat along and across the grain boundary. Because electrons carry the current along the grain boundary, the corresponding hole current is negligible and so are the gradients of hole quasi-Fermi level. These gradients across the grain boundary are *a priori* not negligible. High surface recombination velocities at the grain boundary core can produce strong hole currents transverse to the grain boundary. In this case, our analysis relies on the assumption of high hole mobility, which enables strong hole currents without the need for gradients of hole quasi-Fermi level.

We further assume that the grain boundary charge does not change with applied voltage. This is justified by the limit of large defect density $Q_{\rm GB}(V)/(q\rho_{\rm GB}) \ll 1$. This assumption implies that f, the nonequilibrium counterpart to f_0 in Eq. (5), obeys $f = f_0 \approx 1/2$. This constraint leads to the relation

$$\sum_{j=1}^{M} f_{\rm GB}(E_j) = M/2. \tag{8}$$

Because Eq. (8) depends on the grain boundary carrier densities, the relative sizes of the terms in the occupancy Eq. (2) defines three regimes with distinct properties:

Gaury, Benoit; Haney, Paul.
"Analytical description of charged grain boundary recombination in polycrystalline thin film solar cells."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

- "n-type" grain boundary: In this case, the grain boundary electron density determines the defect occupancy. f remains fixed because of the pinning of the *electron* quasi-Fermi level to the neutral point of the gap states $E_{\rm GB}$ for this case. We also assume that the electron quasi-Fermi level is relatively flat and equal to its bulk value. This is valid because the high electron density in the grain boundary core enables high currents with relatively small quasi-Fermi level gradients.
- "p-type" grain boundary: In this case, the grain boundary hole density determines the occupancy of the defect states. f remains fixed because of the pinning of the *hole* quasi-Fermi level to the neutral point of the gap states $E_{\rm GB}$ for this case. The relatively low electron density at the grain boundary core forces the electron quasi-Fermi level to develop gradients to drive the electron current along the grain boundary. In this case we solve a one-dimensional diffusion equation for the electron density along the grain boundary to obtain the carrier densities and recombination rate.
- High recombination: For sufficiently large applied voltages, the electron and hole carrier densities are the dominating terms in f and determine the defect states occupancies. There is no pinning of quasi-Fermi levels in this case, but we find that the grain boundary carrier densities satisfy

$$p_{\rm GB} = \gamma(V) n_{\rm GB} \tag{9}$$

  where the density ratio γ varies weakly with voltage. In this regime, Eq. (9) together with the assumption of flat hole quasi-Fermi level leads to a one-dimensional drift-diffusion equation for electrons confined to the grain boundary. Solving this equation leads to the carrier densities and recombination. In the case where each occupancy is independent of its gap state energy, one can show that $\gamma = 1$ and the carrier densities must be equal for the system to remain electrically neutral.

A final assumption is that the grains are not fully depleted. For doping densities on the order of 10^{15} cm^{-3}, this requirement implies grain size above 2 μm. For reference, a recent measurement [20] found that the average grain size in CdTe thin films (excluding twin boundaries) was 2.3 μm.

IV. Grain boundary dark recombination current

We provide expressions for the grain boundary dark recombination current. The general expression of the grain boundary dark current density reads

$$J_{\rm GB}(V) = \frac{1}{d}\int_0^{L_{\rm GB}} {\rm d}x\ R_{\rm GB}(x), \tag{10}$$

with $L_{\rm GB}$ the length of the grain boundary. $R_{\rm GB}$ is the Schokley-Read-Hall recombination

$$R_{\rm GB}(x) = \sum_{j=1}^{M} \frac{S_n S_p (n_{\rm GB} p_{\rm GB} - n_i^2)}{S_n(n_{\rm GB} + \bar{n}_j) + S_p(p_{\rm GB} + \bar{p}_j)}, \tag{11}$$

with n_i the intrinsic carrier density. The complexity of many defect states, as opposed to a single one, will be incorporated into effective surface recombination velocities. The physical descriptions of the electron and hole transport are the same as in the single state problem, and the resulting grain boundary dark currents are formally identical to the ones presented in Ref. [18].

In the n-type grain boundary the recombination is determined by holes, which flow from the p-type grain interior into the grain boundary core. Because most of the absorber is p-type, the recombination is uniform along the entire grain boundary. The grain boundary carrier densities read

$$n_{\rm GB}(x) = N_C e^{(-E_g + E_{\rm GB})/k_B T} \tag{12}$$

$$p_{\rm GB}(x) = N_V e^{(-E_{\rm GB} + qV)/k_B T}. \tag{13}$$

Using the fact that $S_n n_{\rm GB}$ is much larger than $S_p p_{\rm GB}$ in the recombination Eq. (11) leads to the grain boundary dark current

$$J_{\rm GB}(V) = \frac{\mathcal{S}_p L_{\rm GB}}{d} N_V e^{(-E_{\rm GB} + qV)/k_B T}, \tag{14}$$

where the effective surface recombination velocity $\mathcal{S}_p$ reads

$$\mathcal{S}_p = \sum_{j=1}^{M} \frac{1}{1 + \frac{\bar{n}_j}{\bar{n}_{\rm GB}} + \frac{S_p \bar{p}_j}{S_n \bar{n}_{\rm GB}}}. \tag{15}$$

A close look at $\mathcal{S}_p$ shows that only states with energies $E_g - E_{\rm GB} \lesssim E \lesssim E_{\rm GB}$ contribute significantly to the recombination (blue region in Fig. 2(a)). The upper limit results from the fact that states above $E_{\rm GB}$ are empty of electrons. The lower limit is the energy at which holes are emitted from the defect state to the valence band faster than electrons relax from the conduction band to the defect state.

The p-type grain boundary has a similar interpretation. The recombination is determined by electrons flowing into the grain boundary core from regions of the grain interior where $n > p$. This corresponds to $x < x_0$ in Fig. 1(a). The

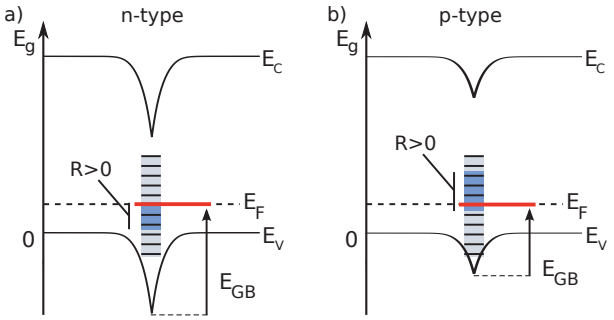


Fig. 2. Schematic of the states contributing to the recombination for the (a) n-type and (b) p-type grain boundary. $E_{\rm GB}$ is the grain boundary neutral energy level. States (not) contributing to the recombination are shaded in blue (grey).

Gaury, Benoit; Haney, Paul.
"Analytical description of charged grain boundary recombination in polycrystalline thin film solar cells."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

recombination is therefore mainly concentrated within the n-region of the pn junction depletion region, and is uniform for $x < x_0$. The grain boundary carrier densities read

$$\begin{aligned} n_{\rm GB}(x) &= N_C e^{(-E_g+E_{\rm GB})/k_BT} e^{qV/k_BT} && \text{for } x < x_0 \\ &= N_C e^{(-E_g+E_{\rm GB})/k_BT} e^{qV/k_BT} e^{-\frac{x-x_0}{L_n}} && \text{for } x > x_0 \end{aligned} \tag{16}$$

$$p_{\rm GB}(x) = N_V e^{-E_{\rm GB}/k_BT}, \tag{17}$$

where $L_n = \sqrt{2D_nL_\mathcal{E}/\mathcal{S}_n}$ (D_n: electron diffusion coefficient) is the electron diffusion length, and $L_\mathcal{E} = k_BT/\mathcal{E}_y$ is the characteristic length of the electric field transverse to the grain boundary $\mathcal{E}_y$. $\mathcal{S}_n$ is the effective surface recombination velocity in this case. Using the fact that $S_pp_{\rm GB}$ is much larger than $S_nn_{\rm GB}$ in the recombination Eq. (11) leads to the grain boundary dark recombination current

$$J_{\rm GB}(V) = \frac{\mathcal{S}_n}{d} N_C e^{(-E_g+E_{\rm GB}+qV)/k_BT} \times \left[x_0 + L_n\left(1 - e^{-\frac{L_{\rm GB}-x_0}{L_n}}\right)\right], \tag{18}$$

where $\mathcal{S}_n$ reads

$$\mathcal{S}_n = \sum_{j=1}^{M} \frac{1}{1 + \frac{\bar{p}_j}{\bar{p}_{\rm GB}} + \frac{S_n\bar{n}_j}{S_p\bar{p}_{\rm GB}}}. \tag{19}$$

Equation (19) shows that only the states with energies $E_{\rm GB} \lesssim E \lesssim E_g - E_{\rm GB}$ contribute significantly to the recombination, as shown in Fig. 2(b). The lower limit results from the fact that states below $E_{\rm GB}$ are empty of holes. The upper limit is the energy at which electrons are emitted from the defect state to the conduction band faster than holes relax from the valence band to the defect state.

As voltage is increased, the grain boundary crosses over to the high-recombination regime. The carrier densities at the grain boundary are now constrained by Eqs (8) and (9). The charge carrier transport is analogous to the single state problem: holes flow towards the pn junction depletion region and the recombination is peaked at a hotspot there. The grain boundary carrier densities read

$$n_{\rm GB}(x) = \frac{1}{\sqrt{\gamma}} n_i e^{qV/(2k_BT)} e^{-\frac{x}{L'_n}} \tag{20}$$

$$p_{\rm GB}(x) = \sqrt{\gamma} n_i e^{qV/(2k_BT)} e^{-\frac{x}{L'_n}}, \tag{21}$$

where $L'_n = \sqrt{4D_nL_\mathcal{E}/\mathcal{S}}$ is the electron diffusion length, and $L_\mathcal{E}$ is the characteristic length of the electric field transverse to the grain boundary. $\mathcal{S}$ is the effective surface recombination velocity in this case

$$\mathcal{S} = \frac{\gamma S_nS_p}{S_n + \gamma S_p} \sum_{j=1}^{M} \frac{1}{1 + \frac{S_n\bar{n}_j + S_p\bar{p}_j}{(S_n+\gamma S_p)\frac{n_i}{\gamma} e^{qV/(2k_BT)}}}. \tag{22}$$

The coefficient γ is found by solving Eq. (8) with the carrier densities Eqs. (20)-(21). The above results lead to the grain boundary dark recombination current

$$J_{\rm GB}(V) = \frac{\mathcal{S}L'_n}{\sqrt{\gamma}d} n_i e^{V/(2V_T)} \left[1 - e^{-L_{\rm GB}/L'_n}\right]. \tag{23}$$

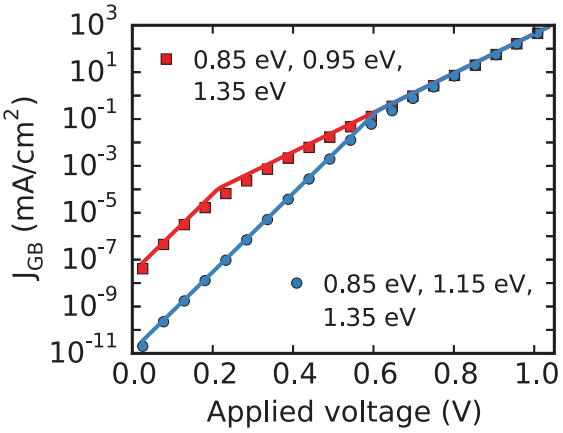


Fig. 3. Grain boundary dark recombination current as a function of voltage for two sets of defects energy levels. Symbols correspond to numerical data, full lines are analytic predictions. Parameters for the simulations are summarized in Table I.

Parameter	Value	Parameter	Value
L	3 μm	$\mu_n = \mu_p$	100 cm^2/(V $\cdot$ s)
d	3 μm	$\tau_n = \tau_p$	10 ns
N_C	8×10^{17} cm^{-3}	$S_{n,p}$	10^5 cm/s
N_V	1.8×10^{19} cm^{-3}	$\rho_{\rm GB}$	10^{14} cm^{-2}
E_g	1.5 eV	ϵ	9.4 ϵ_0
N_A	10^{15} cm^{-3}	N_D	10^{17} cm^{-3}

TABLE I
LIST OF DEFAULT PARAMETERS FOR NUMERICAL SIMULATIONS.

This result differs from the corresponding case in Ref. [18] by the voltage dependence of the effective surface recombination velocity.

To verify the accuracy of these expressions, we solved numerically the drift-diffusion-Poisson set of equations for the geometry presented in Fig. 1(a), and various distributions of defect states. Table I gives a list of the material parameters used for these calculations. We used periodic boundary conditions in the y-direction, and infinite (zero) surface recombination velocity for majority (minority) carriers at the contacts to simulate perfectly selective contacts. Fig. 3 shows the grain boundary recombination current for three gap states. The red (blue) curve corresponds to an n-type (p-type) grain boundary. Because we consider the limit of high defect density of states, the neutral point $E_{\rm GB}$ of each distribution of gap states corresponds to the intermediate energy of each case. A good agreement can be seen between our analytical model and the full numerical simulations.

V. Open-circuit voltage

Upon comparing the electron diffusion length to the length of the grain boundary in the above results, we find that limiting cases of Eqs. (14), (18) and (23) are of the general form

$$J_{\rm GB}(V) = \lambda \frac{S}{d} N e^{-E_a/k_BT} e^{qV/(nk_BT)}, \tag{24}$$

Gaury, Benoit; Haney, Paul.
"Analytical description of charged grain boundary recombination in polycrystalline thin film solar cells."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

where S is a surface recombination velocity, λ is a length characteristic of the recombination region, N is an effective density of states, E_a is an activation energy, n is the ideality factor and V is the applied voltage.

Equation (24) allows us to derive a closed form relation for the open-circuit voltage. It was shown in Ref. [18] that, around $V_{\rm oc}$, the current-voltage relation of an illuminated junction is given by the sum of the short-circuit current $J_{\rm sc}$ and the dark current $J_{\rm dark}(V)$. $V_{\rm oc}$ therefore satisfies $J_{\rm dark}(V_{\rm oc}) = J_{\rm sc}$. Assuming large values of surface recombination velocities, the grain boundary recombination dominates over the bulk recombination and $J_{\rm dark} = J_{\rm GB}$. Solving for $J_{\rm GB}(V_{\rm oc}) = J_{\rm sc}$ therefore leads to the general form of the open-circuit voltage

$$qV_{\rm oc}^{\rm GB} = nE_a + nk_BT\ln\left(\frac{dJ_{\rm sc}}{S\lambda N}\right). \tag{25}$$

Equation (25) provides insight into how the parameters controlling the grain boundary and the distribution of gap states affect $V_{\rm oc}$. For instance, the open-circuit voltage increases linearly with the defect activation energy. This energy corresponds to the neutral point of the distribution of gap states in the n-type and p-type regimes. However, $E_a = E_g/2$ in the high-recombination regime, so that the impact the initial distribution of gap states on $V_{\rm oc}$ is minimal (it remains present in the effective surface recombination velocity).

VI. Conclusion

We derived analytical relations for the grain boundary dark recombination current and the open-circuit voltage of a pn^+ junction with a positively charged columnar grain boundary. These new expressions account for a distribution of defect states in the gap of the absorber material. We showed that our simplified analytical model describes the essential features of the full numerical calculations.

Acknowledgment

B. G. acknowledges support under the Cooperative Research Agreement between the University of Maryland and the National Institute of Standards and Technology Center for Nanoscale Science and Technology, Award 70NANB14H209, through the University of Maryland.

References

[1] S. G. Kumar and K. K. Rao, "Physics and chemistry of cdte/cds thin film heterojunction photovoltaic devices: fundamental and critical aspects," *Energy Environ. Sci.*, vol. 7, no. 1, pp. 45–102, 2014.

[2] M. A. Green, K. Emery, Y. Hishikawa, W. Warta, E. D. Dunlop, D. H. Levi, and A. W. Y. Ho-Baillie, "Solar cell efficiency tables (version 49)," *Prog. Photovolt. Res. Appl.*, vol. 25, no. 1, pp. 3–13, 2017.

[3] P. M. Haney, H. P. Yoon, B. Gaury, and N. B. Zhitenev, "Depletion region surface effects in electron beam induced current measurements," *J. Appl. Phys.*, vol. 120, p. 095702, 2016.

[4] H. P. Yoon, P. M. Haney, D. Ruzmetov, H. Xu, M. S. Leite, B. H. Hamadani, A. A. Talin, and N. B. Zhitenev, "Local electrical characterization of cadmium telluride solar cells using low-energy electron beam," *Sol. Energ. Mat. Sol. Cells*, vol. 117, pp. 499–504, 2013.

[5] O. Zywitzki, T. Modes, H. Morgner, C. Metzner, B. Siepchen, B. Späth, C. Drost, V. Krishnakumar, and S. Frauenstein, "Effect of chlorine activation treatment on electron beam induced current signal distribution of cadmium telluride thin film solar cells," *J. Appl. Phys.*, vol. 114, no. 16, p. 163518, 2013.

[6] C. Li, Y. Wu, J. Poplawsky, T. J. Pennycook, N. Paudel, W. Yin, S. J. Haigh, M. P. Oxley, A. R. Lupini, M. Al-Jassim, S. J. Pennycook, and Y. Yan, "Grain-boundary-enhanced carrier collection in cdte solar cells," *Phys. Rev. Lett.*, vol. 112, no. 15, p. 156103, 2014.

[7] I. Visoly-Fisher, S. R. Cohen, and D. Cahen, "Direct evidence for grain-boundary depletion in polycrystalline cdte from nanoscale-resolved measurements," *Appl. Phys. Lett.*, vol. 82, no. 4, pp. 556–558, 2003.

[8] H. Moutinho, R. Dhere, C.-S. Jiang, Y. Yan, D. Albin, and M. Al-Jassim, "Investigation of potential and electric field profiles in cross sections of CdTe/CdS solar cells using scanning kelvin probe microscopy," *J. Appl. Phys.*, vol. 108, no. 7, p. 074503, 2010.

[9] C.-S. Jiang, R. Noufi, J. AbuShama, K. Ramanathan, H. Moutinho, J. Pankow, and M. Al-Jassim, "Local built-in potential on grain boundary of Cu(In, Ga)Se_2 thin films," *Appl. Phys. Lett.*, vol. 84, no. 18, pp. 3477–3479, 2004.

[10] H. C. Card and E. S. Yang, "Electronic processes at grain boundaries in polycrystalline semiconductors under optical illumination," *IEEE Trans. Electron Devices*, vol. 24, no. 4, pp. 397–402, April 1977.

[11] J. G. Fossum and F. A. Lindholm, "Theory of grain-boundary and intragrain recombination currents in polysilicon pn-junction solar cells," *IEEE Trans. Electron Devices*, vol. 27, no. 4, pp. 692–700, 1980.

[12] M. A. Green, "Bounds upon grain boundary effects in minority carrier semiconductor devices: A rigorous perturbationapproach with application to silicon solar cells," *J. Appl. Phys.*, vol. 80, no. 3, pp. 1515–1521, 1996.

[13] S. Edmiston, G. Heiser, A. Sproul, and M. Green, "Improved modeling of grain boundary recombination in bulk and p-n junction regions of polycrystalline silicon solar cells," *J. Appl. Phys.*, vol. 80, no. 12, pp. 6783–6795, 1996.

[14] M. Gloeckler, J. R. Sites, and W. K. Metzger, "Grain-boundary recombination in Cu(In,Ga)Se_2 solar cells," *J. Appl. Phys.*, vol. 98, no. 11, p. 113704, 2005.

[15] K. Taretto and U. Rau, "Numerical simulation of carrier collection and recombination at grain boundaries in Cu(In, Ga)Se_2 solar cells," *J. Appl. Phys.*, vol. 103, no. 9, p. 094523, 2008.

[16] U. Rau, K. Taretto, and S. Siebentritt, "Grain boundaries in Cu(In, Ga)(Se, S)$_2$ thin-film solar cells," *Appl. Phys. A*, vol. 96, no. 1, pp. 221–234, 2009.

[17] F. Troni, R. Menozzi, E. Colegrove, and C. Buurma, "Simulation of current transport in polycrystalline CdTe solar cells," *J. Electron. Mater.*, vol. 42, no. 11, pp. 3175–3180, 2013.

[18] B. Gaury and P. M. Haney, "Charged grain boundaries reduce the open-circuit voltage of polycrystalline solar cells–An analytical description," *J. Appl. Phys.*, vol. 120, p. 234503, 2016.

[19] B. Gaury and P. M. Haney, "Anatomy of charged grain boundaries in polycrystalline solar cells," *arXiv preprint arXiv:1704.04234*, 2017.

[20] J. Moseley, W. K. Metzger, H. R. Moutinho, N. Paudel, H. L. Guthrey, Y. Yan, R. K. Ahrenkiel, and M. M. Al-Jassim, "Recombination by grain-boundary type in cdte," *J. Appl. Phys.*, vol. 118, no. 2, 2015.

Gaury, Benoit; Haney, Paul.
"Analytical description of charged grain boundary recombination in polycrystalline thin film solar cells."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

Analytic description of the impact of grain boundaries on Voc

Paul Haney[1] and Benoit Gaury[1,2]

[1]Center for Nanoscale Science and Technology National Institute of Standards and Technology Gaithersburg, MD, 20899, USA
[2]Maryland NanoCenter, University of Maryland College Park, MD 20742

***Abstract*—The impact of grain boundaries on the performance of polycrsytalline photovoltaics remains an open question. We present a simplified description of dark grain boundary recombination current. The dark current takes the form of a diode equation, and the model provides closed form expressions for the reverse saturation current and ideality factor in terms of grain boundary and system parameters. This model applies under conditions relevant for thin film photovoltaics such as CdTe, namely for p-type absorbers with reasonably high bulk hole mobility, positively charged grain boundaries with high defect density, and grains which are not fully depleted. The dark recombination current can be used to predict the open circuit voltage for a given short circuit density, providing a simple closed form expression which shows how grain boundaries impact $V_{\rm oc}$.**

***Index Terms*—polycrystalline solar cell, grain boundary recombination, open circuit voltage**

I. Introduction

Thin films photovoltaics like CdTe and Cu(In,Ga)Se_2 exhibit high conversion efficiency despite their high defect density [1]. Grain boundaries are a primary source of defects, however so far polycrystalline samples outperform their single crystal counterparts [2]. This apparent dichotomoy between the electrical and structural properties invites the unexpected question: "Can grain boundaries be beneficial for photovoltaic performance?" There is not a clear consensus on this issue. In our view there is not a universal answer to this question, it depends on the details of the grain boundary and bulk properties. Grain boundaries may improve the performance of samples with exceedingly poor bulk properties, but are always detrimental to samples with reasonably good bulk properties. In the best cases, grain boundaries assist carrier collection at zero bias, increasing the short circuit density $J_{\rm sc}$. This is due to the built-in electric field accompanying charged grain boundaries which separates carriers and inhibits recombination (provided that carriers driven to the grain boundary are majority carriers at the grain boundary core) [3], [4]. However in *all* cases grain boundaries are harmful for the open circuit voltage $V_{\rm oc}$ [5], [6]. This is not surprising since $V_{\rm oc}$ is reduced by recombination, and recombination is enhanced by grain boundaries when the device is under forward bias. However an intuitive, quantitative description of how grain boundaries reduce $V_{\rm oc}$ is still lacking, despite previous numerical and analytical works [5], [6], [7]. This article describes our recently developed models which provide this simple relation between grain boundary properties and $V_{\rm oc}$.

II. Grain boundary recombination

In a recent set of papers [8], [9], we have presented analytical expressions for the grain boundary dark recombination current. These expressions take the form of a general diode equation:

$$J_{\rm GB}(V) = J_0 \exp\left(\frac{qV}{nk_BT}\right), \tag{1}$$

where $q > 0$ is the electron charge, V is the applied voltage, $k_{\rm B}$ is the Boltzmann constant, and T is the temperature. The result of our work is closed form expressions for the reverse saturation current J_0 and ideality factor n in terms of grain boundary and system parameters. We also demonstrated that the dark recombination current can be used to accurately predict the open circuit voltage for a given short circuit current density. In particular, numerical simulations indicate that the following relation holds:

$$V_{\rm oc} = \frac{nk_BT}{q} \ln\left(\frac{J_{\rm sc}}{J_0}\right). \tag{2}$$

The impact of grain boundary recombination on $V_{\rm oc}$ can therefore be concisely quantified. This may enable rational approaches to mitigating the negative consequences of grain boundary recombination.

Our analysis showed that the system behavior depends on the grain boundary defect type (*e.g.* donor or acceptor), the location of the defect energy level(s) with respect to midgap, the applied voltage, and other factors. The system behavior can be classified into several regimes, where each regime has its own peculiarities and requires its own detailed analysis. The myriad of different cases can obscure the overall picture of grain boundary recombination. Despite the differences between different regimes, a single framework for understanding the system response can be presented, which offers useful perspective when viewing the multiplicity of cases. In this work we aim to provide a more global, qualitative description of our model of grain boundary recombination current.

Before discussing grain boundary recombination, it's useful to rewrite Eq. (1) in a form which helps to frame our analysis. We write the dark recombination current as:

$$J = N\left(\frac{\lambda}{\tau}\right) \exp\left(\frac{qV}{nk_BT}\right). \tag{3}$$

Haney, Paul; Gaury, Benoit.
"Analytic description of the impact of grain boundaries on Voc."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

The form of Eq. (3) is fully general for thermally activated transport. Current density can always be written as the product of a density N times a velocity: length λ divided by time τ. The exponential factor describes the classic diode voltage dependence with ideality factor n. Although Eq. (3) is generic, when applied to a 1-dimensional p-n junction, the factors in Eq. (3) acquire quite specific physical interpretations. N and τ correspond to the density and effective lifetime of the species controlling the recombination, and λ is the length scale over which recombination occurs. The ideality factor n is determined by the nature of recombination. An ideality factor $n = 1$ corresponds to minority carrier-controlled recombination, while $n = 2$ applies when both species contribute to recombination.

We first demonstrate this interpretation of Eq. (3) by evaluating the well-known 1-dimensional p-n junction dark recombination current. For this system, dark recombination current is the sum of two contributions: the diffusion current and junction recombination current. Diffusion current is associated with minority carrier recombination in the neutral region. For concreteness, we suppose the system is a pn^+ junction and consider recombination in the p-type region. Electrons are minority carriers, so N is given by the equilibrium electron density in the p-type neutral region, which we denote n_p^0. The lifetime is set by the bulk electron lifetime, $\tau = \tau^e_{\mathrm{bulk}}$. Minority carriers undergo simple diffusion in the neutral region, so the length scale of the minority carrier density profile is the diffusion length, $\lambda = \sqrt{D^e_{\mathrm{bulk}}\tau^e_{\mathrm{bulk}}}$, where D^e_{bulk} is the bulk electron diffusivity. Recombination is determined by the minority carrier (electron) density, so the ideality factor $n = 1$. Substituting these factors for Eq. (3) reproduces the well-known result for diffusive dark current. We next apply the same analysis to junction recombination current. In this case the recombination occurs in the depletion region and involves both nonequilibrium electrons and holes. The equilibrium density of electrons and holes in the depletion region is given by $N = n_i$, where n_i is the intrinsic density. The length scale over which recombination occurs is set by the depletion width W of the pn junction, $\lambda \approx W$. The lifetime is determined by bulk recombination, $\tau = \tau_{\mathrm{bulk}}$. Both nonequilibrium electrons and holes participate in recombination, so that $n = 2$. These factors correctly reproduce the junction recombination dark current.

For grain boundary recombination, the same interpretations of the parameters entering Eq. (3) apply. What remains is to identify the recombination species and its lifetime, and to determine the region over which recombination occurs. Doing so requires knowledge of how carriers behave in the 2-dimensional model, which we discuss next.

As in many analytical models, the key aspect of our treatment lies in the approximations we make. The initial problem in its fully general form is not analytically tractable, due to nonlinearities and its 2 (or 3)-dimensional nature. Reducing the problem to a more manageable form requires valid, simplifying assumptions, which in turn require sufficient knowledge of system behavior. The rough picture is that positively charged grain boundaries lead to downward band bending in most of the p-type absorber, providing a confinement potential for electrons. We assume one end of the grain boundary is in close proximity with the n-contact, so electrons are efficiently funnelled into the grain boundary and the vast majority of electron current is carried along the grain boundary core. Hole current takes place in the grain bulk, and is directed towards regions of high recombination. We assume grains are not fully depleted. In practice this corresponds to grain sizes which exceed $1\ \mu\mathrm{m}$ in CdTe. Hole current therefore requires very small gradients in the hole quasi-Fermi level. This fact leads to a key assumption: that the hole quasi-Fermi level E^{Fp} is approximately flat everywhere. This assumption is valid only for sufficiently high intragrain hole mobility; for typical material parameter values of CdTe, the intragrain hole mobility μ_p should exceed $30\ \mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$ (see ref. [8] for details of this estimate).

We restrict our attention to the recombination at the grain boundary core, reducing the domain of interest to one dimension. However by itself this does not simplify the problem: The solution at the grain boundary depends on the solution in the grain interior, and the problem remains essentially 2-dimensional. The assumption of flat E^{Fp} is crucial here, as it implies $E^{Fp}_{\mathrm{GB}} = E^{Fp}_{\mathrm{bulk}}$. This relation provides a link between the grain boundary and grain bulk and enables the analysis of the two domains to be separated. The dimensionality of the problem is then reduced from 2 to 1.

Our next assumption is that the grain boundary defect density is large. The charge of the grain boundary defect is equal to the defect density multiplied by a statistical factor related to the defect occupation and type (donor or acceptor, see Eq. (7)). If the defect density is very large, the statistical factor must be very small to ensure that the defect charge remains finite. A large defect density also implies that the statistical factor is independent of applied voltage [9]. This leads to a constraint which takes the place of the Poisson equation. The implications of this constraint depend on the details of the grain boundary, such as the position of the defect level with respect to midgap. This is the point at which different cases and their analysis bifurcates. We discuss these cases later. For now the salient point is that this assumption of high defect density provides another simplification to the problem.

Having made these assumptions, the problem can be reduced to a single 1-dimensional effective diffusion equation for the electron quasi-Fermi level $E^{Fn}_{\mathrm{GB}}(x)$, where x is the coordinate along the grain boundary core. It's not surprising that a description of grain boundary recombination would include a continuity equation for E^{Fn}_{GB}: electrons are confined to the grain boundary and carry the recombination current there [5]. The 1-dimensional diffusive motion of electrons along the grain boundary is parameterized by the electron grain boundary diffusivity D^e_{GB} and effective lifetime τ^e_{GB}. These parameters can be expected to differ from their bulk counterparts due to the highly defective grain boundary core.

Haney, Paul; Gaury, Benoit.
"Analytic description of the impact of grain boundaries on Voc."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

TABLE I
RECOMBINATION CURRENT PARAMETERS

n-type	p-type	high recombination
$N = p^0_{\mathrm{GB}}$	$N = n^0_{\mathrm{GB}}$	$N = n_i$
$\lambda = L_{\mathrm{GB}}$	$\lambda = x_0 + L^{e\ \mathrm{diff}}_{\mathrm{GB}}$	$\lambda = L^{e\ \mathrm{diff}}_{\mathrm{GB}}$
$S = S_p^{\mathrm{eff}}$	$S = S_n^{\mathrm{eff}}$	$S = \sqrt{S_n^{\mathrm{eff}} S_p^{\mathrm{eff}}}$
$n = 1$	$n = 1$	$n = 2$

The grain boundary-confined electron diffusivity D^e_{GB} is likely reduced from the bulk value due to increased disorder scattering. D^e_{GB} enters as a free parameter in the model. The electron effective lifetime is reduced from its bulk value due to proximity to the grain boundary recombination centers. The recombination strength of the grain boundary is parameterized by an effective surface recombination velocity for electrons (holes) $S^{\mathrm{eff}}_{n(p)}$. The effective lifetime τ^e_{GB} is determined by the length scale of electron confinement L_E according to $\tau^e_{\mathrm{GB}} = L_E/S^{\mathrm{eff}}_n$. The confinement length L_E is determined by the built-in potential around the grain boundary, which is associated with two length scales: the depletion width of the grain boundary and $k_{\mathrm{B}}T/\left(qE_{\mathrm{GB}}\right)$, where E_{GB} is the magnitude of the electric field at the grain boundary. The appropriate choice for L_E depends on the regime of system behavior, but either choice gives qualitatively similar results. The diffusion length of electrons confined to the grain boundary L^e_{GB} is then given by $L^e_{\mathrm{GB}} = \sqrt{D^e_{\mathrm{GB}}\tau^e_{\mathrm{GB}}} = \sqrt{D^e_{\mathrm{GB}}L_E/S^{\mathrm{eff}}_n}$.

With this picture of the system in mind, we turn again to Eq. (3) and specify the factors entering the formula. The appropriate lifetime τ for Eq. (3) is S^{eff}/d, where d is the grain size (the appropriate S^{eff} depends on the grain boundary type). This amounts to taking all of the recombination centers concentrated on the grain boundary surface and smearing them out uniformly across the entire grain. This is the crudest approximation one can make to account for grain boundary recombination, and has been used rather successfully to describe polycrystalline Si solar cells [11]. For photovoltaics with high grain boundary defect density, simply re-scaling τ and applying 1-dimensional p-n junction theory is inadequate for determining N, λ, and n. For these three parameters, the analysis proceeds differently according to the majority carrier type at the grain boundary core. As always, defect-mediated recombination is controlled by the minority carrier density. The minority carrier type depends on the Fermi level at the grain boundary core, which in turn depends on grain boundary defect properties (mostly the defect energy level).

Beginning with an n-type grain boundary core, holes are minority carriers, so N is the equilibrium grain boundary hole density, which we denote by p^0_{GB}. Since recombination is set by only one species, the ideality factor $n = 1$. The relevant S is that for holes: $S = S^{\mathrm{eff}}_p$. We consider a single donor+acceptor defect, in which case the effective surface recombination velocity is equal to the bare surface recombination velocity: $S^{\mathrm{eff}}_p = S_p$. (For other cases like the single donor defect and the continuum of donor+acceptor defects, the effective surface recombination velocity differs

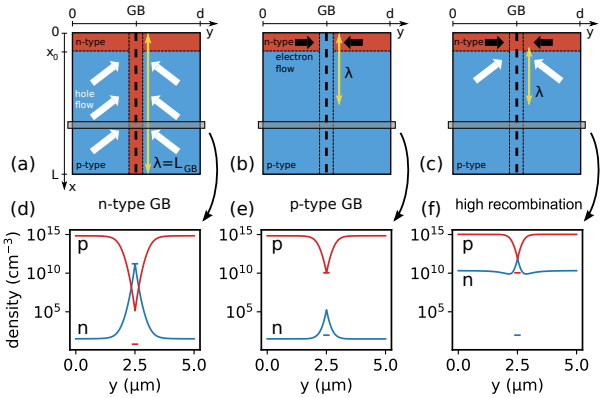


Fig. 1. (a), (b), and (c) show the system geometry and the electron and hole current flow for n-type, p-type, and high recombination grain boundaries, respectively. Also shown is the length λ over which recombination occurs along the grain boundary core. The grain boundary is columnar and positioned at $y = d/2$. (d), (e), and (f) show the electron and hole density under forward bias through a slice of the neutral region perpendicular to the grain boundary for n-type, p-type, and high recombination grain boundaries, respectively. The grain boundary core is located at $y = 2.5$ μm, and the blue (red) tick marks at the grain bounday core position indicate the values of $\overline{n}_{\mathrm{GB}}$ ($\overline{p}_{\mathrm{GB}}$).

from the bare surface recombination velocity.) λ is determined by the region over which holes are available to flow into the grain boundary and recombine. The majority of the grain boundary length L_{GB} is embedded in p-type bulk, so holes are available for transport into the grain boundary from the grain bulk over approximately the entire grain boundary. Hence $\lambda = L_{\mathrm{GB}}$ (see Fig. 1(a)).

For a p-type grain boundary, similar reasoning immediately leads to $N = n^0_{\mathrm{GB}}$, $n = 1$, and $S = S_n$. The recombination length λ includes the region over which electrons are available to flow into the grain boundary. We delimit this region with x_0: for $x < x_0$, $n_{\mathrm{bulk}} > p_{\mathrm{bulk}}$. Electrons flow into the grain boundary for $x < x_0$ and propagate via diffusion along the grain boundary core for $x > x_0$, where additional recombination occurs (see Fig. 1(b)). The total length over which recombination occurs is therefore $x_0 + L^{e\ \mathrm{diff}}_{\mathrm{GB}}$, where $L^{e\ \mathrm{diff}}_{\mathrm{GB}}$ is the diffusion length of electrons confined near the grain boundary core. Based on the discussion of electron diffusion along the grain boundary given earlier, we have $\lambda = x_0 + L^{e\ \mathrm{diff}}_{\mathrm{GB}} = x_0 + \sqrt{D^e_{\mathrm{GB}}L_E/S^{\mathrm{eff}}_n}$.

In the case where electron and hole density are of comparable magnitude, we immediately anticipate that $N = n_i$ and $n = 2$. To determine λ, we note that both electrons and holes must be transported to the grain boundary for recombination (see Fig. 1(c)). Holes are available along almost the entire length of the grain boundary, as discussed earlier. However electrons are only available for $x < x_0$. λ is therefore set by the electrons' availability. Recombination is concentrated near the middle of the depletion region, and decays on a length scale of the electron diffusion length, so that $\lambda = L^{e\ \mathrm{diff}}_{\mathrm{GB}}$. These cases are summarized in Table 1, and depicted in Fig. 1 (a)-(c).

Haney, Paul; Gaury, Benoit.
"Analytic description of the impact of grain boundaries on Voc."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

Having summarized our understanding of grain boundary recombination in qualitative terms, we next place it in more quantitative context by presenting the relevant governing equations. The key underlying equations provide the grain boundary occupation f^{GB}, grain boundary charge Q^{GB}, and grain boundary recombination R^{GB} in terms of carrier densities and the defect properties. These are all standard equations which can be found in textbooks [12], [13]. We start with the grain boundary occupancy:

$$f^{\mathrm{GB}} = \frac{S_n n_{\mathrm{GB}} + S_p \overline{p}_{\mathrm{GB}}}{S_n \left(n_{\mathrm{GB}} + \overline{n}_{\mathrm{GB}}\right) + S_p \left(p_{\mathrm{GB}} + \overline{p}_{\mathrm{GB}}\right)}. \tag{4}$$

In the above, n_{GB} and p_{GB} are the electron and hole carrier density at the grain boundary, respectively (note that n_{GB}, p_{GB}, and f^{GB} can vary as a function of position along the grain boundary core). $\overline{n}_{\mathrm{GB}}$ and $\overline{p}_{\mathrm{GB}}$ are the electron and hole density one would obtain if the Fermi level is positioned at the defect energy level E_{GB}. If E_{GB} is measured from the valence band, then:

$$\overline{n}_{\mathrm{GB}} = N_c \exp\left[\left(E_{\mathrm{GB}} - E_g\right)/k_{\mathrm{B}}T\right], \tag{5}$$

$$\overline{p}_{\mathrm{GB}} = N_v \exp\left[-E_{\mathrm{GB}}/k_{\mathrm{B}}T\right]. \tag{6}$$

The charge of the grain boundary is given by:

$$Q^{\mathrm{GB}} = q\rho^{\mathrm{GB}} \times \begin{cases} \left(1 - f^{\mathrm{GB}}\right), & (\text{donor}) \\ \left(-f^{\mathrm{GB}}\right), & (\text{acceptor}) \\ \left(1 - 2f^{\mathrm{GB}}\right) & (\text{donor} + \text{acceptor}) \end{cases} \tag{7}$$

Here ρ^{GB} is the two-dimensional defect density at the grain boundary core. Note that the charge of the defect state depends on its type (donor or acceptor). Experimental evidence indicates positively charged grain boundaries [10], so we restrict our attention to the donor and donor+acceptor cases (the acceptor case only yields negatively charged grain boundaries). Note that for the donor+acceptor case, both defects are assumed to be present at the same energy E_{GB}. The donor and acceptor defects therefore compensate each other only when the Fermi energy is equal to the defects' common energy level.

The grain boundary charge Q^{GB} is compensated by the surrounding depletion region charge. For a depletion width W, the associated space charge is $2qN_AW$, where N_A is the doping and the factor of 2 arises from having depletion regions on both sides of the grain boundary. The relation $Q^{\mathrm{GB}} = 2qN_AW$ determines the equilibrium band bending between grain boundary and grain bulk.

The assumption we described earlier is that ρ^{GB} is "large". The factors in parentheses in Eq. (7) must therefore be "small" for Q^{GB} to remain finite. For the donor case, this implies $\left(1 - f^{\mathrm{GB}}\right)$ is small, or $f_{\mathrm{GB}} \approx 1$. For the donor+acceptor case, we have $f^{\mathrm{GB}} \approx 1/2$: both donor and acceptor state are approximately half-occupied. As described in Ref. [9], large ρ^{GB} also implies the the factor in parentheses does not change with applied voltage; we can determine f^{GB} in equilibrium and use the same value under forward bias.

Having determined f^{GB}, we can return to Eq. (4) and consider what values of carrier density n_{GB}, p_{GB} are needed to yield the correct value of f^{GB}. We'll consider the donor+acceptor case here as an example, in which $f^{\mathrm{GB}} = 1/2$. In equilibrium, $f^{\mathrm{GB}} = 1/2$ is satisfied by the carrier densities $n_{\mathrm{GB}} = \overline{n}_{\mathrm{GB}}$ and $p_{\mathrm{GB}} = \overline{p}_{\mathrm{GB}}$. This means that the Fermi energy is pinned to E_{GB} (see discussion preceding Eqs. (5)-(6)). In equilibrium, the majority carrier type at the grain boundary core is therefore determined solely by E_{GB}; if E_{GB} is closer to the conduction (valence) band, the grain boundary core is n (p)-type.

Let's consider an n-type grain boundary under forward bias. The terms n_{GB} and $\overline{n}_{\mathrm{GB}}$ are much larger than p_{GB} and $\overline{p}_{\mathrm{GB}}$ in the expression on the right-hand-side of Eq. (4). An applied bias voltage induces nonequilibrium electron and hole densities. Provided p_{GB} is much less than n_{GB}, n_{GB} must remain fixed to the value $\overline{n}_{\mathrm{GB}}$ to maintain $f_{\mathrm{GB}} = 1/2$. This is demonstrated in Fig. 1(d), which shows the electron and hole density under forward bias through a slice of the system perpendicular to the grain boundary core. The small blue (red) tick mark at $y = 2.5$ μm indicates the value of $\overline{n}_{\mathrm{GB}}$ ($\overline{p}_{\mathrm{GB}}$). In this plot, $n_{\mathrm{GB}} = \overline{n}_{\mathrm{GB}}$ while p_{GB} exceeds $\overline{p}_{\mathrm{GB}}$. We can write the grain boundary hole density in terms of the bulk hole density using the assumption we described earlier: $E^{Fp}_{\mathrm{GB}} = E^{Fp}_{\mathrm{bulk}}$. This leads to $p_{\mathrm{GB}} = p^0_{\mathrm{GB}} \exp\left(qV/k_{\mathrm{B}}T\right)$. The minority carrier density increases exponentially with applied voltage. The same behavior holds for p-type grain boundaries, where majority and minority carrier types are interchanged with respect to the n-type grain boundary (see Fig. 1(e)).

Having determined the majority and minority carrier density, we can compute the grain boundary recombination R_{GB}, given by:

$$R_{\mathrm{GB}} = \frac{S_n S_p \left(n_{\mathrm{GB}} n_{\mathrm{GB}} - n_i^2\right)}{S_n \left(n_{\mathrm{GB}} + \overline{n}_{\mathrm{GB}}\right) + S_p \left(p_{\mathrm{GB}} + \overline{p}_{\mathrm{GB}}\right)} \tag{8}$$

For the n-type grain boundary, Eq. (8) simplifies to $R_{\mathrm{GB}} = S_p p^0_{\mathrm{GB}} \exp\left(qV/k_{\mathrm{B}}T\right)$. We argued previously that the length scale over which recombination occurs is equal to the entire grain boundary length L_{GB}. Together these factors reproduce the diode equation terms for an n-type grain boundary as given in Table 1.

As the applied voltage increases, the minority carrier density p_{GB} dutifully increases exponentially and will eventually approach n_{GB}. At this point further increases in the applied voltage push both n_{GB} and p_{GB} to exceed $\overline{n}_{\mathrm{GB}}$. Now $f_{\mathrm{GB}} = 1/2$ is satisfied by insisting that $S_n n_{\mathrm{GB}} = S_p p_{\mathrm{GB}}$ (see Eq. (4)). This is the high-recombination regime. Returning to Eq. (8) and utilizing the relation $n_{\mathrm{GB}} p_{\mathrm{GB}} = n_i^2 \exp\left(qV/k_{\mathrm{B}}T\right)$, the maximum value of recombination along the grain boundary core can be found as:

$$R^{\max}_{\mathrm{GB}} = \sqrt{S_n S_p}\, n_i \exp\left(\frac{qV}{2k_{\mathrm{B}}T}\right) \tag{9}$$

In the high recombination regime, both nonequilibrium electrons and holes control the recombination. These carriers are transported to the grain boundary from the bulk. The

Haney, Paul; Gaury, Benoit.
"Analytic description of the impact of grain boundaries on Voc."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

point of maximum grain boundary recombination occurs in the depletion region. In this regime an electrostatic potential gradient is also developed along the grain boundary which drives the high electron current. We showed that the electron and hole density profiles decay along the grain boundary over a length scale of $L_{\mathrm{GB}}^{e\ \mathrm{diff}}$ in this case [8]. This leads to the recombination which decays away from its maximum value with the same length scale, as shown in Fig. 1(c) and the 3rd column of Table 1. This ends the analysis of the donor+acceptor case.

The analysis of the donor cases proceeds along similar lines. One important wrinkle is that $f_{\mathrm{GB}} \approx 1$ in this case (see discussion following Eq. (7)). The defect is nearly fully occupied, which implies that for the same value of E_{GB}, the band bending is larger in the donor case than in the donor+acceptor case. This in turn implies the grain boundary hole density is suppressed in the donor case relative to the donor+acceptor case. For this reason, the effective hole recombination velocity is given by $S_p^{\mathrm{eff}} = (1 - f_{\mathrm{GB}})\, S_p$. $(1 - f_{\mathrm{GB}})$ is small by assumption, so the hole-controlled recombination is significantly reduced in the donor defect case [9].

III. Conclusion

We end the overview here, and refer the reader to references for a fuller and more rigorous derivation of the grain boundary dark current, as well as demonstrations that the grain boundary dark current can be used to predict V_{oc}. The type of analysis can be extended to consider a continuum of defect states, grain boundaries with arbitrary orientation, and networks of inhomogeneous grain boundaries. With this degree of model flexibility, we can begin to consider the behavior of realistic, inhomogeneous polycrystalline materials.

Acknowledgment

B. G. acknowledges support under the Cooperative Research Agreement between the University of Maryland and the National Institute of Standards and Technology Center for Nanoscale Science and Technology, Award 70NANB14H209, through the University of Maryland.

References

[1] R. M. Geisthardt, M. Topi Marko J. R. and Sites, "Status and potential of CdTe solar-cell efficiency", *IEEE Journal of photovoltaics* vol. 5, pp. 1217-1221, 2015.

[2] I. Visoly-Fisher, S. R. Cohen, A. Ruzin, D. Cahen, "Polycrystalline Devices Can Outperform Single-Crystal Ones: Thin Film CdTe/CdS Solar Cells", *Advanced Materials* vol. 16, pp. 879883, 2004.

[3] C. Li, Y. Wu, J. Poplawsky, T. J. Pennycook, N. Paudel, W. Yin, S. J. Haigh, M. P. Oxley, A. R. Lupini, M. Al-Jassim, and others, "Grain-boundary-enhanced carrier collection in CdTe solar cells", *Phys. Rev. Lett.*, vol. 112, pp. 156103, 2014.

[4] M. Tuteja, P. Koirala, V. Palekis, S. MacLaren, C. S. Ferekides, R. W. Collins, and A. A. Rockett, "Direct Observation of CdCl2 Treatment Induced Grain Boundary Carrier Depletion in CdTe Solar Cells Using Scanning Probe Microwave Reflectivity Based Capacitance Measurements", *J. Phys. Chem. C*, vol. 120, pp. 7020-7024, 2016.

[5] M. Gloeckler, J. R. Sites, and W. K. Metzger, "Grain-boundary recombination in $\mathrm{Cu(In,Ga)Se_2}$ solar cells", *J. App. Phys.*, vol. 98, pp. 113704, 2005.

[6] K. Taretto and U. Rau, "Numerical simulation of carrier collection and recombination at grain boundaries in $\mathrm{Cu(In,Ga)Se_2}$ solar cells", *J. App. Phys.*, vol. 103, pp. 094523, 2008.

[7] S. A. Edmiston, G. Heiser, A. B. Sproul, and M. A. Green, "Improved modeling of grain boundary recombination in bulk and p-n junction regions of polycrystalline silicon solar cells", *J. App. Phys.*, vol. 80, pp. 6783-6795, 1996.

[8] B. Gaury, and P. M. Haney, "Charged grain boundaries reduce the open-circuit voltage of polycrystalline solar cellsAn analytical description", *J. App. Phys.*, vol. 120, pp. 234503, 2016.

[9] B. Gaury and P. M. Haney, "Anatomy of charged grain boundaries in polycrystalline solar cells", *arXiv preprint arXiv:1704.04234*, 2017.

[10] H. P. Yoon, P. M. Haney, D. Ruzmetov, H. Xu, M. S. Leite, B. Hamadani, A. A. Talin, N. B. Zhitenev, "Local electrical characterization of cadmium telluride solar cells using low-energy electron beam", *Sol. En. Mat. and Sol. Cells*, vol. 117, pp. 499-504, 2013.

[11] H. C. Card, and E. S. Yang, Edward S, "Electronic processes at grain boundaries in polycrystalline semiconductors under optical illumination", *IEEE Transactions on Electron Devices*, vol. 24, pp. 397-402, 1977.

[12] S. J. Fonash, *Solar cell device physics*: Academic Press, 1981.

[13] E. S. Yang, *Fundamentals of semiconductor devices*: McGraw-Hill Companies, 1978.

Haney, Paul; Gaury, Benoit.
"Analytic description of the impact of grain boundaries on Voc."
Paper presented at 44th IEEE Photovoltaic Specialists Conference (PVSC 2017), Washington, D.C., United States. June 25, 2017 - June 30, 2017.

Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)

Irena Bojanova; Paul E. Black
NIST, Gaithersburg, USA
irena.bojanova@nist.gov, paul.black@nist.gov

Yaacov Yesha
NIST, Gaithersburg, USA; UMBC, Baltimore, USA
yaacov.yesha@nist.gov

***Abstract*—Accurate, precise, and unambiguous definitions of software weaknesses (bugs) and clear descriptions of software vulnerabilities are vital for building the foundations of cybersecurity. The Bugs Framework (BF) comprises rigorous definitions and (static) attributes of bug classes, along with their related dynamic properties, such as proximate, secondary and tertiary causes, consequences, and sites. This paper presents an overview of previously developed BF classes and the new cryptography related classes: Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN). We analyze corresponding vulnerabilities and provide their clear descriptions by applying the BF taxonomy. We also discuss the lessons learned and share our plans for expanding BF.**

Keywords—software weaknesses; bug taxonomy; attacks.

I. Introduction

Advances in scientific foundations of cybersecurity rely on the availability of accurate, precise, and unambiguous definitions of software weaknesses (bugs) and clear descriptions of software vulnerabilities. The myriad unprecedented attacks and security exposures, including on Internet of Things (IoT) applications, calls for serious efforts towards such formalization.

To provide a foundation, we are developing the Bugs Framework (BF) [1], which organizes bugs into distinct classes, such as buffer overflow (BOF), injection (INJ), faulty operation (FOP), and control of interaction frequency bugs (CIF). Each BF class has an accurate and precise definition and comprises: level (added after [1]), causes, attributes, consequences, and sites of bugs. Closely related classes may be grouped in clusters. *Level* (high or low) identifies the fault as language-related or semantic. *Causes* bring about the fault. At least one *attribute* (denoted as underlined) identifies the software fault, while the rest may be simply descriptive. It is useful to catalog possible *consequences* of faults. *Sites* are locations in code (identifiable mainly for low level classes) where the bug might occur under circumstances indicated by the causes. The goal of BF is to help researchers and practitioners more accurately and quickly diagnose, describe, and measure security vulnerabilities.

In this paper, we summarize the BF classes we previously developed, then detail our newly-developed cryptography-related classes: Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN). The details include definitions and taxonomy of these classes, examples of vulnerabilities from the Common Vulnerabilities and Exposures (CVE) [2], and corresponding Common Weakness Enumerations (CWE) [3] or Software Fault Patterns (SFP) [4]. The final section summarizes our work, discusses lessons we learned, and presents our plans for expanding BF further.

II. Previously Developed BF Classes

Our first developed BF classes were: Buffer Overflow (BOF), Injection (INJ), and Control of Interaction Frequency Bugs (CIF) [1]. Here we only give their definitions and attributes. Details and examples of use are available at https://samate.nist.gov/BF/.

BOF: *The software accesses through an array a memory location that is outside the boundaries of that array.* Attributes: Access, Boundary, Location, Magnitude, Data Size, Reach.

INJ: *Due to input with language-specific special elements, the software assembles a command string that is parsed into an invalid construct.* Attributes: Invalid Construct, Language, Special Element, Entry Point.

CIF: *The software does not properly limit the number of repeating interactions per specified unit.* Attributes: Interaction, Number, Unit, Actor.

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

III. Cryptographic Store or Transfer Bugs

A. *Cryptography*

Cryptography is a broad, complex, and subtle area. It incorporates many clearly separate processes, such as encryption/decryption, verification of data or source, and key management. There are bugs if the software does not properly transform data into unintelligible form, verify authenticity or correctness, manage keys, or perform other related operations. Some transformations require keys, for example encryption and decryption, while others do not, for example secret sharing. Authenticity covers integrity of data, identity of data source, origin for non-repudiation, and content of secret sharing. Correctness is verified for uses such as zero-knowledge proofs. Cryptographic processes use particular algorithms to achieve particular security services [5].

Examples of attacks are spoofing messages, brute force attack, replaying instructions, timing attack, chosen plaintext attack, chosen ciphertext attack, and exploiting use of weak or insecure keys.

In this paper, we use cryptographic store or transfer to illustrate our ENC, VRF, and KMN classes of bugs. Note that these classes may appear in many other situations such as self-sovereign identities [6], block ciphers, and threshold cryptography. We focus on transfer (or store) because it is well known and it is what most people think of when "cryptography" is mentioned. We define bugs in cryptographic store or transfer as: The software does not properly encrypt/decrypt, verify, or manage keys for data to be securely stored or transferred.

B. *Our Model*

A modern, secure, flexible cryptographic storage or transfer protocol likely involves subtle interaction between encryption, verification, and key management processes. It may involve multiple stages of agreeing on encryption algorithms, establishing public and private keys, creating session keys, and digitally signing texts for verification. Thus, encryption may use key management, which

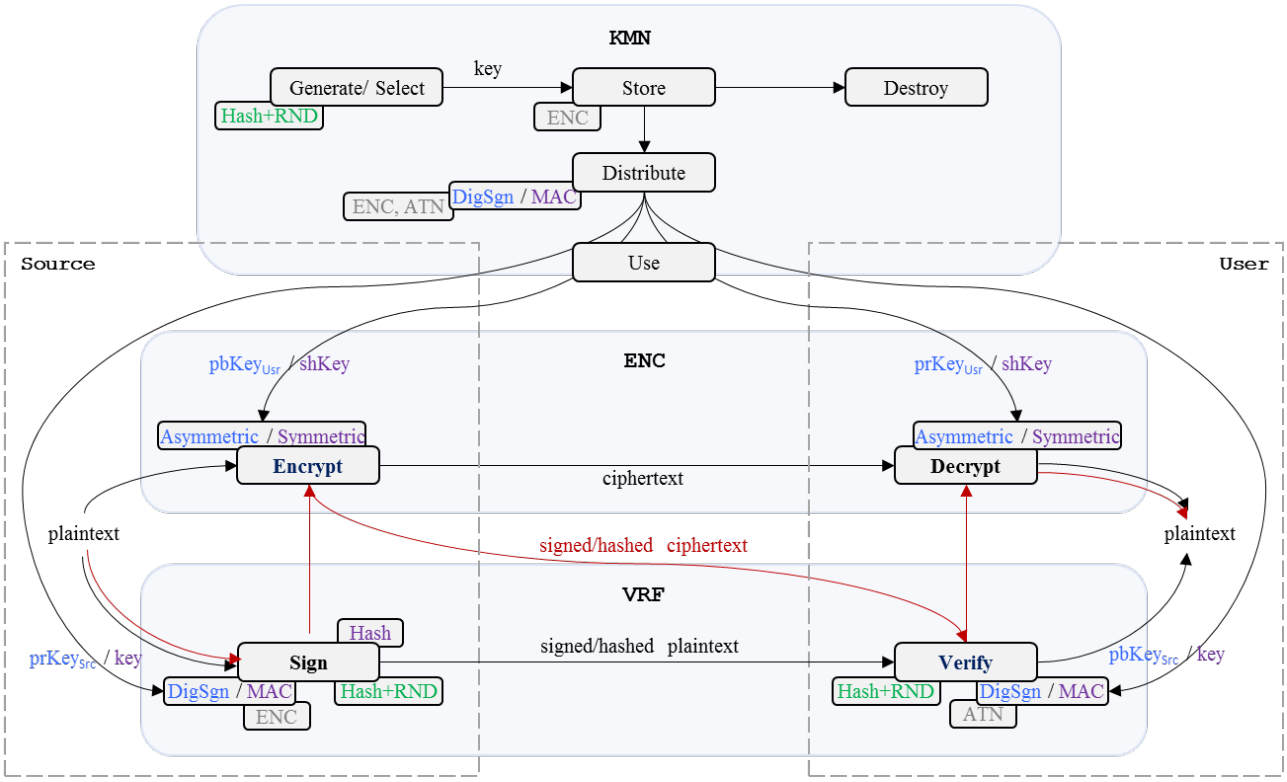


Fig. 1. Our Model of Cryptographic Store or Transfer Bugs. Encryption may occur in tandem with Verification or it may precede Verification serially, if the cipertext is signed or hashed. Encryption uses Key Management, and Key Management likely uses Encryption and Verification to handle keys.

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

itself uses encryption and verification. Fig. 1 presents a model of these recursive interactions and where potentially the corresponding ENC, VRF, KMN, and other BF bugs could happen. The rounded rectangles indicate the boundaries of the classes. The dashed ones show sending and receiving entities.

KMN is a class of bugs related to key management. Key management comprises key generation, selection, storage, retrieval and distribution, and determining and signaling when keys should be abandoned or replaced. A particular protocol may use any or all of these operations. Key management could be by a third party, the source, or the user – thus the KMN area intersects the Source and User areas. A third-party certificate authority (CA) distributes public keys in signed certificates. Key management often uses a recursive round of encryption and decryption, and verification to establish a shared secret key or session key before the actual plaintext is handled.

ENC is a class of bugs related to encryption and decryption. Encryption is by the source, decryption is by the user. The encryption/decryption algorithm may be symmetric, that is uses the same key for both, or asymmetric, which uses a pair of keys, one to encrypt and the other to decrypt. Public key cryptosystems are asymmetric. Ciphertext may be sent directly to the user, and verification accompanies it separately. The red line is a case where plaintext is signed or hashed and then encrypted.

VRF is a class of bugs related to verification. Verification takes a key and either plaintext or ciphertext, signs or hashes it, then passes the result to the user. The user uses the same key or the other key from the key-pair to verify data integrity or source. Note that hash alone without any other mechanism cannot be used to verify source or to protect data integrity against attackers. However, it can be used to protect data integrity against channel errors [5].

In the cases of symmetric encryption, one secretly shared key (shKey) is used. The source encrypts with shKey, and the user decrypts also with shKey. In the cases of asymmetric encryption, pairs of mathematically related keys are used. The source pair is $pbKey_{Src}$ and $prKey_{Src}$; the user pair is $pbKey_{Usr}$ and $prKey_{Usr}$. The source encrypts with $pbKey_{Usr}$ and signs with $prKey_{Src}$. The user decrypts with $prKey_{Usr}$ and verifies with $pbKey_{Src}$.

IV. Encryption/Decryption Bugs Class – ENC

A. Definition

We define Encryption Bugs (ENC) as:

The software does not properly transform sensitive data (plaintext) into unintelligible form (ciphertext) using cryptographic algorithm and key(s).

We define also Decryption Bugs as:

The software does not properly transform ciphertext into plaintext using cryptographic algorithm and key(s).

Note that "transform" is for confidentiality.

ENC is related to KMN, Randomization (RND), and Information Exposure (IEX).

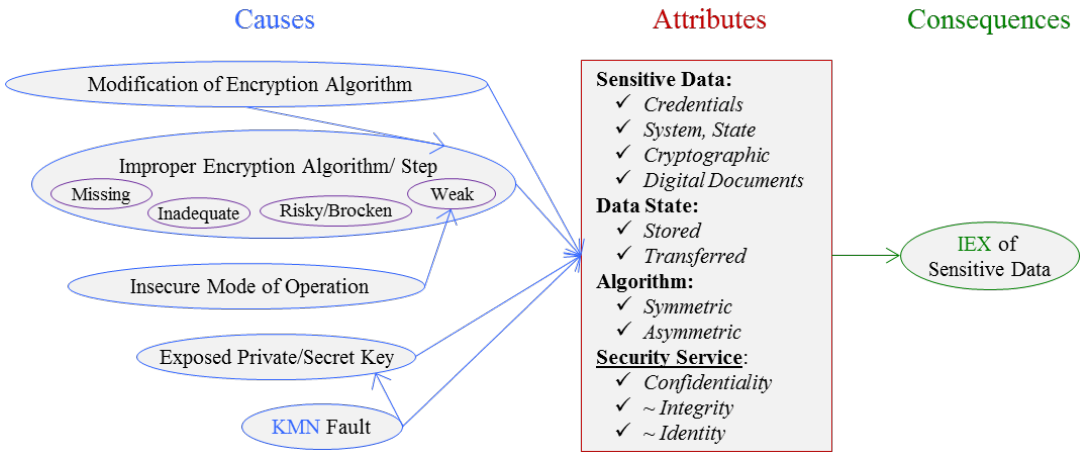


Fig. 2. The Encryption Bugs (ENC) class represented as causes, attributes and consequences.

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

B. Taxonomy

Fig. 2 depicts ENC causes, attributes and consequences. In the graph of causes, modification of algorithm is remove/change or add a cryptographic step. Improper algorithm or step could be missing, inadequate, weak, risky/broken. Insecure mode of operation leads to weak encryption algorithm.

The attributes of ENC are:

Sensitive Data – Credentials, System Data, State Data, Cryptographic Data, Digital Documents. This is secret (confidential) data. Credentials include password, token, smart card, digital certificate, biometrics (fingerprint, hand configuration, retina, iris, voice.) System Data could be configurations, logs, Web usage. Cryptographic Data is hashes, keys, and other keying material.

Data State – Stored, Transferred. This reflects if data is in rest or use, or if data is in transit. Secure store is needed for data that is in rest or use from files (e.g. ini, temp, configuration, log server, debug, cleanup, email attachment, login buffer, executable, backup, core dump, access control list, private data index), directories (Web root, FTP root, CVS repository), registry, cookies, source code & comments, GUI, environmental variables. Secure transfer is needed also for data in transit between processes or over a network.

Algorithm – Symmetric, Asymmetric. This is the key encryption scheme used to securely store/transfer sensitive data. Symmetric (secret) key algorithms (e.g. Serpent, Blowfish) use one shared key. Asymmetric (public) key algorithms (e.g. Diffie-Hellman, RSA) use a pair of keys: public and private.

Security Service(s) – Confidentiality (and Integrity and Identity Authentication). This is the security service that was failed by the encryption process. Confidentiality is the main security service provided by encryption. Those marked with '~' are only for some specific modes of encryption.

ENC is a high level class, so sites do not apply.

C. Examples

1) CVE-2002-1946

This vulnerability is listed in [7] and discussed in [8, 9, 10]. Our BF description is:

ENC: *Use of weak symmetric encryption algorithm (one-to-one mapping) allows confidentiality failure of stored (in registry) sensitive data (passwords), which may be exploited for IEX of that sensitive data (passwords).*

Analysis (based on [8, 9, 10]): The one-to-one mapping uses two fixed arrays of characters. There was no remedy as of 09/01/2014!

2) CVE-2002-1697

This vulnerability is listed in [11] and discussed in [12, 13, 14]. Our BF description is:

ENC: *Use of insecure mode of operation (ECB) leads to weak symmetric encryption algorithm (for same shared key produces same ciphertext from same plaintext) and allows confidentiality failure of transferred sensitive data, which may be exploited for IEX of that sensitive data.*

Analysis (based on [12, 13, 14]): Using electronic codebook (ECB) results in weak encryption, that produces the same ciphertext from the same plaintext blocks. This is a case of deterministic encryption, where patterns in plaintext become evident in the ciphertext.

D. Related CWEs and SFP

CWEs related to ENC are: CWE-256, 257, 261, 311-318, 325, 326, 327, 329, 780 [3].

The related SFP clusters are SPF 17.1 Broken Cryptography and SFP 17.2 Weak Cryptography under Primary Cluster: Cryptography [4]. Note that some of the CWEs listed there are not ENC.

V. Verification Bugs Class – VRF

A. Definition

We define Verification Bugs (VRF) as:

The software does not properly sign data, check and prove source, or assure data is not altered.

Note that "check" is for identity authentication, "prove" is for origin (signer) non-repudiation, and "not altered" is for integrity authentication.

VRF is related to KMN, RND, ENC, Authentication (ATN), IEX.

B. Taxonomy

Fig. 3 depicts VRF causes, attributes and consequences. In the graph of causes, modification of algorithm and improper algorithm or step have the same meaning as in ENC.

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

The attributes of VRF are:

Verified Data – Secret, Public. This is the data that needs verification. It may be confidential or public. Secret (confidential) data could be cryptographic hashes, secret keys, or keying material. Public data could be signed contract, documents, or public keys.

Algorithm – Hash Function + RND, message authentication code (MAC), Digital Signature. Hash functions are used for integrity authentication. They may use RND. MAC are symmetric key algorithms (one secret key per source/user), used for integrity authentication, identity authentication. It needs authentication code generation, source signs data, user gets tag for key and data, and verifies data by tag and key. Digital Signature is an asymmetric key algorithm (two keys), used for integrity and identity authentication, and origin (signer) non-repudiation. It needs key generation, signature generation, and signature verification. MAC and Digital Signature use KMN and recursively VRF.

Security Service – Data Integrity Authentication, Identity Authentication, Origin (Signer) Non-Repudiation. This is the security service the verification process failed. Integrity Authentication is for data and keys. Identity Authentication and Origin Non-Repudiation are for source authentication.

VRF is a high level class, so sites do not apply.

C. *Examples*

1) CVE 2001-1585

This vulnerability is listed in [15] and discussed in [16, 17]. Our BF description is:

VRF: *Missing verification step (challenge-response of private key using digital signature) in public key authentication allows identity authentication failure, which may be exploited for IEX.*

Analysis (based on [16, 17]): The step that should be included is challenge-response authentication: The client is required by the server to sign a message using the client's private key. Successful verification of that signature by the server, using the public key, confirms that the client owns the private key that is paired with that public key, and therefore that client should be allowed to login. That challenge-response authentication step is missing.

2) CVE-2015-2141

This vulnerability is listed in [18] and discussed in [19, 20, 21]. Our BF description is:

VRF: *Modification of digital signature verification algorithm (Rabin-Williams) by adding a step (blinding) leads to recovery of private key, that allows identity authentication failure, which may be exploited for IEX.*

Analysis (based on [19, 20, 21]): Having the private key allows an attacker to be authenticated as the owner of that key.

The software intends to use blinding to defend against a timing attack, as follows: Instead of signing the data directly, the data is first transformed using a secret random value (blinding) and then is digitally signed using a private key. At the end, the effect is removed (unblinding), so that there is signed data as if no transformation took place. See [20, 21] for blinding used for RSA.

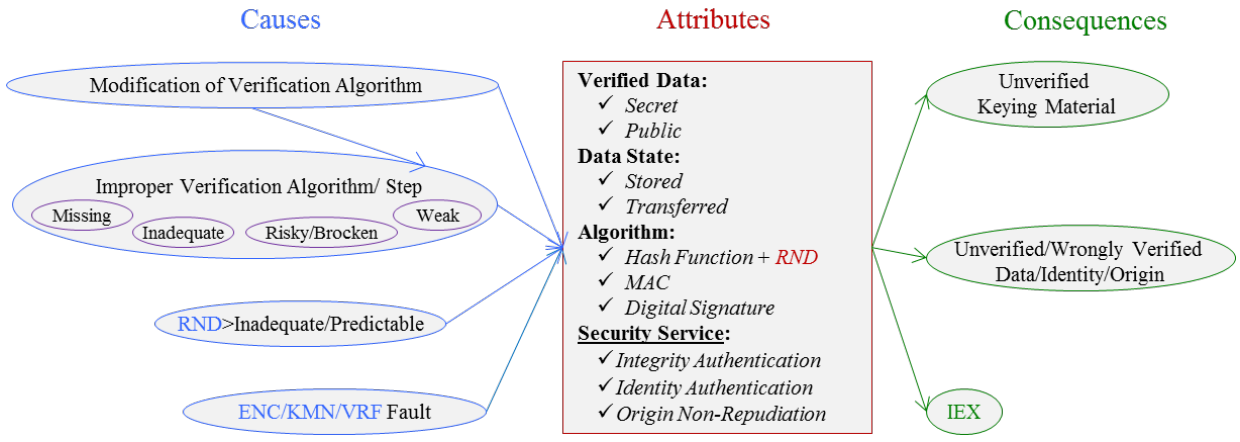


Fig. 3. The Verification Bugs (VRF) class represented as causes, attributes and consequences.

The flaw in this CVE is in doing blinding/ unblinding incorrectly, so that in some cases the effect of the transformation is not removed from the data. This enables the attacker to use the transformed data to recover the private key using a mathematical calculation as described in [20]. In [20] it is observed that if the secret random integer used to transform the message is a quadratic residue modulo an appropriate integer, then the unblinding step correctly undoes the transformation. The fix in [20] assures that the integer is such a quadratic residue.

D. *Related CWEs and SFP*

CWEs related to VRF are: CWE-295-296, 347 [3].

The related SFP cluster is 17.2 Weak Cryptography under Primary Cluster: Cryptography [4]. Note that some of the listed CWEs are not VRF.

VI. Key Management Bugs Class – KMN

A. *Definition*

We define Key Management Bugs (KMN) as:

The software does not properly generate, store, distribute, use, or destroy cryptographic keys and other keying material.

KMN is related to ENC, RND, VRF, IEX.

B. *Taxonomy*

Fig. 4 depicts KMN causes, attributes and consequences.

The attributes of KMN are:

Cryptographic Data – Hashes, Keying Material, Digital Certificate.

Algorithm – Hash Function + RND, MAC, Digital Signature. Different cryptosystem have their own key generation algorithm(s).

Operation – Generate or Select, Store, Distribute, Use, Destroy. This is the failed operation. Generate uses RND. Store includes update and recover. Distribute includes key establishment, transport, agreement, wrapping, encapsulation, derivation, confirmation, shared secret creation; uses ENC and KMN (reclusively).

KMN is a high level class, so sites do not apply.

C. *Examples*

1) *CVE-2016-1919*

This vulnerability is listed in [22] and discussed in [23]. Our BF description is:

A KMN leads to an ENC.

KMN: *Use of weak algorithm (eCryptFS-key from password and stored TIMA key) allows generation of keying material (secret key) that can be obtained through brute force attack, which may be exploited for IEX of keying material (the secret key).*

ENC: *KMN fault leads to exposed secret key that allows confidentiality failure of stored sensitive data, which may be exploited for IEX of that sensitive data.*

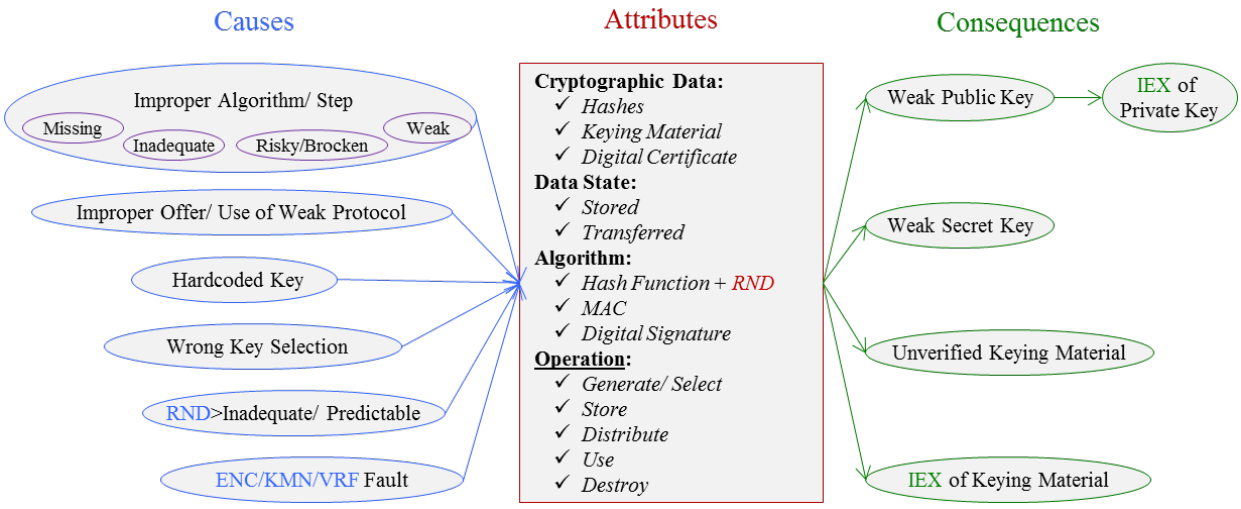


Fig. 4. The Key Management Bugs (KMN) class represented as causes, attributes and consequences

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

Analysis (based on [23]): The set of possible keys is a known small set.

The TIMA key is a random stored byte string. The secret key used is obtained by XOR of the TIMA key and the password characters, where the minimum password length is 7. However, if the password length is no more than 8, a base 64 expansion results in a key that does not depend on the password. The TIMA key is stored, and for a known TIMA key, the key is known, or, if the password length slightly exceeds 8, there is a small set of possible keys. The TIMA key can be obtained using a preliminary step.

2) CVE-2015-0204, 1637, 1067 (FREAK - Factoring attack on RSA-ExportKeys)

This vulnerability is listed in [24, 25, 26] and discussed in [27, 28, 29, 30]. Our BF description is:

An inner KMN leads to an inner ENC, which leads to an outer ENC.

Inner KMN: *Client-accepted improper offer of weak protocol (SSL with Export RSA) from MITM-tricked server allows use of an algorithm (Export RSA) that generates 512-bit keying material (pair of keys), for which private key may be obtained through factorization of public key, which may be exploited for IEX of keying material (the private key).*

Inner ENC: *KMN fault leads to exposed private key for asymmetric encryption (RSA) that allows confidentiality failure of transferred sensitive data (Pre-Master Secret), which may be exploited for IEX of sensitive data (Master Secret).*

Outer ENC: *KMN fault leads to exposed secret key (Master Secret) for symmetric encryption allows confidentiality failure of transferred sensitive data (passwords, credit cards, etc.), which may be exploited for IEX of that sensitive data (passwords, credit cards, etc.).*

Inner KMN and inner ENC only set up the secret key. Outer ENC is the actual general data transfer.

Interestingly in this example the consequence from the first bug (inner KMN) causes the second bug (inner ENC), whose consequences cause the third bug (outer ENC). The inner KMN is a server bug, sending a weak key, (that the client did not ask for), intended for KMN use by client (encrypting Pre-Master Secret). It is also a client bug, as the client accepted the offer of using the insecure method, and therefore the server proceeded. The client could have refused that offer. The inner ENC is a client bug, using that weak key to encrypt the Pre-Master Secret, and then transmitting that weakly encrypted Pre-Master Secret over a network that is not secure.

Analysis (based on [27, 28, 29, 30]): The server offers a weak protocol (Export RSA) while the client requested strong protocol (RSA).

Communication is encrypted by symmetric encryption. The key for that encryption (Master Secret) is created by both client and server from a Pre-Master Secret and nonces sent by client and server. The Pre-Master Secret is sent encrypted by RSA cryptosystem. The client requests RSA protocol, but man in the middle (MITM) intercepts and requests Export RSA that uses a 512 bit key. Factoring a 512 bit RSA key is feasible.

Because of a bug, the client agrees to Export RSA. MITM factors the public 512 bit public RSA key, uses this factoring to recover the private RSA key, and then uses that private key to decrypt the Pre-Master Secret. Then it uses the Pre-Master Secret and the nonces to generate the Master Secret. The Master Secret enables MITM to decrypt the encrypted communication from that point on.

D. Related CWEs and SFP

CWEs related to KMN are: CWE-321, 322, 323, 324 [3].

The related SFP clusters are SFP 17.2 Weak Cryptography under Primary Cluster: Cryptography and SFP 4.13 Digital Certificate under Primary Cluster: Authentication [4]. Note that, some of the CWEs listed in 17.2 are not KMN.

VII. Concluding Remarks

A. Summary

We presented three new BF classes: Encryption/Decryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN). They join other rigorously-defined classes: Injection (INJ), Control of Interaction Frequency Bugs (CIF), and Buffer Overflow (BOF). We presented the (static) attributes of the classes, along with the classes' causes and consequences.

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

We analyzed particular vulnerabilities related to those classes and provided clear descriptions. We showed that the BF-structured description of FREAK, using KMN and ENC, is quite concise and still far clearer than unstructured explanations that we have found.

B. *Lessons Learned*

At first, we tried to define a Cryptography class with attributes, causes, and consequences. However, we realized that subsidiary processes, like randomization and verification, are used in completely different contexts. As we developed a model of cryptographic store and transfer bugs, we learned how rich and subtle the relations between processes were. We ended up defining separate ENC, VRF, and KMN classes of bugs.

We found that a model of the cryptographic store or transfer processes helps: it illustrated the relation between the classes and helped us determine what operations should be included and what should not. We learned that the structure of BF is *not* settled. It may need to be refined further.

C. *Future Work*

One of our next steps is to explain more cryptographic bugs using ENC, VRF, and KMN. We also need to explore the use of ENC, VRF, and KMN in contexts other than cryptographic store and transfer. This will show where BF structures need refinements.

Another step is to develop other BF classes. We are currently working on Randomization Bugs (RND), Authentication Bugs (ATN), Authorization Bugs (AUT), Information Exposure (IEX), Faulty Operation (FOP), and Memory Allocation Bugs (MAL). FOP includes faults during arithmetic operations, such as integer overflow and divide by zero. MAL includes memory allocation faults, like use after free, multiple free, and failure to free.

Our goal is BF to become the software developers' and testers' "Best Friend."

REFERENCES

[1] I. Bojanova, P. E. Black, Y. Yesha, and Y. Wu, "The Bugs Framework (BF): A Structured Approach to Express Bugs," 2016 IEEE International Conference on Software Quality, Reliability, and Security (QRS 2016). Vienna, Austria. August 1-3 2016.

[2] The MITRE Corporation, CVE Common Vulnerabilities and Exposures, (CVE), http://www.cve.mitre.org.

[3] The MITRE Corporation, Common Weakness Enumeration (CWE), http://cwe.mitre.org.

[4] N. Mansourov, "DoD Software Fault Patterns," KDM Analytics, Inc., 2011. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADB381215.

[5] E. Barker, "NIST Special Publication 800-175B Guideline for Using Cryptographic Standards in the Federal Government: Cryptographic Mechanisms," August 2016. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-175B.pdf.

[6] A. Tobin and D. Reed, "The Inevitable Rise of Self-Sovereign Identity", September 2016, https://sovrin.org/wp-content/uploads/2017/07/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf.

[7] The MITRE Corporation, CVE-2002-1946, http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2002-1946.

[8] The MITRE Corporation, CWE 326, https://cwe.mitre.org/data/definitions/326.html.

[9] SecurityTracker. VSNL Integrated Dialer Weak Encoding Discloses Passwords to Local Users Alert ID: 1005515, http://securitytracker.com/id/1005515.

[10] IBM X-Force Exchange, Integrated Dialer Software stores passwords using weak encryption algorithm: CVE-2002-1946, https://exchange.xforce.ibmcloud.com/vulnerabilities/10517.

[11] The MITRE Corporation, CVE-2002-1697, http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2002-1697.

[12] Wikipedia, RSA (cryptosystem), http://en.wikipedia.org/wiki/RSA_(cryptosystem).

[13] Seclists, Security weaknesses of VTun, http://seclists.org/bugtraq/2002/Jan/119.

[14] Wikipedia, Deterministic encryption, https://en.wikipedia.org/wiki/Deterministic_encryption.

[15] The MITRE Corporation, CVE-2001-1585, http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2001-1585.

[16] OpenBSD Security Advisory, Authentication By-Pass Vulnerability in OpenSSH-2.3.1, http://www.openbsd.org/advisories/ssh_bypass.txt.

[17] S. Tatham, PuTTY User Manual – Chapter 8: "Using public keys for SSH authentication," http://the.earth.li/~sgtatham/putty/0.60/htmldoc/Chapter8.html.

[18] The MITRE Corporation, CVE-2141, http://www.cve.mitre.org/cgi-bin/cvename.cgi?name=2015-2141

[19] Bugzilla – Bug 936435, VUL-0: CVE-2015-2141: libcryptopp: libcrypto++ -- security update, https://bugzilla.suse.com/show_bug.cgi?id=936435.

[20] E. Sidorov, "Breaking the Rabin-Williams digital signature system implementation in the Crypto++ library," 2015, http://eprint.iacr.org/2015/368.pdf.

[21] Wikipedia, "Blinding Cryptography," https://en.wikipedia.org/wiki/Blinding_(cryptography).

[22] The MITRE Corporation, CVE-2016-1919, https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2016-1919.

[23] openwall.net, [CVE-2016-1919] "Weak eCryptFS Key generation from user password," http://lists.openwall.net/bugtraq/2016/01/17/2.

[24] The MITRE Corporation, CVE--2015-0204, https://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2015-0204.

[25] The MITRE Corporation, CVE--2015-1637, https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-1637.

[26] The MITRE Corporation, CVE--2015-1067, https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-1067.

[27] R. Heaton, The SSL FREAK vulnerability explained, http://robertheaton.com/2015/04/06/the-ssl-freak-vulnerability.

[28] Censys, The FREAK Attack. https://censys.io/blog/freak

[29] StackExchange, Protecting phone from the FREAK bug, http://android.stackexchange.com/questions/101929/protecting-phone-from-the-freak-bug/101966.

[30] GitHub, openssl, Only allow ephemeral RSA keys in export ciphersuites, https://github.com/openssl/openssl/commit/ce325c60c74b0fa784f5872404b722e120e5cab0?diff=split.

Bojanova, Irena; Black, Paul; Yesha, Yaacov.
"Cryptography Classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)."
Paper presented at 2017 IEEE 28th Annual Software Technology Conference (STC), Gaithersburg, MD, United States. September 25, 2017 - September 28, 2017.

INVESTIGATING ALTERATION OF PRE-VIKING HILLFORT GLASSES FROM THE BROBORG HILLFORT SITE, SWEDEN

Carolyn I. Pearce[1], Jamie L. Weaver[2], Edward P. Vicenzi[3], Thomas Lam[3], Paula DePriest[3], Robert Koestler[3], Tamas Varga[1], Micah D. Miller[1], Bruce W. Arey[1], Michele A. Conroy[1], John S. McCloy[4], Rolf Sjoblom[5], Michael J, Schweiger[1], David K. Peeler[1], and Albert A. Kruger[6]

1. Pacific Northwest National Laboratory, Richland, WA USA; 2. National Institute of Standards and Technology, Gaithersburg, MD USA; 3. Museum Conservation Institute, Smithsonian Institution, Suitland, MD USA; 4. School of Mechanical and Materials Engineering, Washington State University, Pullman, WA USA; 5. Luleå University of Technology, Luleå, Sweden; 6. Department of Energy, Office of River Protection, Richland, WA USA

Keywords

Nuclear waste glass disposal, glass corrosion, long-term analogues, biodeterioration

Introduction

Radioactive waste, produced as a result of nuclear fission for power generation and weapons production, must be immobilized to limit radionuclide release into the biosphere over periods of many thousands of years. Several countries, including the United States, have chosen to vitrify nuclear waste materials prior to disposal. The solubility of ions in liquid, including if the liquid is super-cooled to a glass phase, is greater than in a corresponding crystalline solid; thus, radioactive ions are incorporated into a broad distribution of available sites in the glass structure, and the product is highly durable [1]. However, a glass phase is not thermodynamically stable and will, in principle, undergo some degree of alteration with time. Low level vitrified wastes will be disposed of in near surface sites, such as the Integrated Disposal Facility at the Hanford Nuclear Reservation, WA. Near-field solution chemistry, water diffusion, ion exchange, precipitation of mineral alteration phases, and microbial colonization could influence long-term glass performance, but not all of these parameters are currently captured in the modeling and performance assessments for low level nuclear waste glass disposal sites. Thus, robust models based on a mechanistic understanding of the processes responsible for glass degradation and radionuclide release in near surface environments are needed. These models will give confidence to performance assessments for nuclear waste disposal sites by predicting the durability of vitrified nuclear wastes over the course of thousands of years.

Alteration rates of nuclear waste glasses have been predominantly determined by short-term (a few days to a few months) laboratory alteration tests. These tests provide information on initial rate(s) of corrosion, occurring at a fast rate but only over a short period of time. Per one model of glass alteration, the first step in alteration (Stage I) is initiated by hydrolysis of the silicate network, followed by rapid dissolution and ion exchange between the glass and its altering medium. The rate of glass dissolution then decreases by several orders of magnitude as the gel-like alteration layer, formed in Stage I, is stabilized (Stage II). This results in a leveling out of the dissolution rate and a pseudo equilibrium between formation and dissolution of the alteration layer. Occasionally, given the correct chemical conditions, glass alteration rates have been observed to increase by orders of magnitude (Stage III), concomitant with the precipitation and growth of crystalline secondary phases (zeolites, magnesium silicates, iron silicates, etc.). Short-term experiments are incompletely informative with regard to predicting how a glass will alter in

Weaver, Jamie; Pearce, Carolyn; Vicenzi, Edward; Lam, Thomas; DePriest, Paula; Koestler, Robert; Varga, Tamas; Miller, Micah; Arey, Bruce; Conroy, Michele; McCloy, John; Sjoblom, Rolf; Schweiger, Michael; Peeler, David; Kruger, Albert.
"Investigating Alteration of Pre-Viking Hillfort Glasses From the Broborg Hillfort Site, Sweden."
Paper presented at Materials Science & Technology 2017 (MS&T 17), Pittsburgh, PA, United States. October 8, 2017 - October 12, 2017.

the long-term (Stage II and III). Experts in the field [2, 3, 4] maintain that studying analogues which: (i) are representative of long time scales; (ii) have similar chemical characteristics to the waste glass; and (iii) have been altered under comparable conditions to those expected for nuclear waste disposal, can indeed be used to address this issue [5]. For a good analogue, the altering environment over time and the original chemistry of the unaltered glass, must be known sufficiently well [4, 5]. These analogues can be used to validate glass alteration models, currently based on short-term ($\leq \approx 10$ yrs) tests with real or simulated nuclear waste glass, by comparing the alteration processes and products observed to those predicted by the models.

Methods

The anthropogenic glass from vitrified hillforts, specifically from the Broborg hillfort, are viable analogues for nuclear waste glass [6, 7]. Broborg is a ~1500 year old hillfort located near Uppsala, Sweden. Broborg hillfort glasses have a wide range of chemistries, including dark (basalt-like) glass and clear (silica-rich) glass [8]. Vitrification of the walls at Broborg has been ascribed to melting of rock rich in amphibole [7, 9]. Information on the excavation of the Broborg glass sample analyzed in this study, along with a description of the sample chemistry, can be found in [8, 10-12]. The sample is of historical significance, and special handling and sampling procedures were implemented so that only small portions of glass were extracted in a manner that was not detrimental to structural integrity of the object. The sample was analyzed using: (i) X-ray computed tomography (XCT) to assess internal microstructure and define key regions of interest for sectioning; (ii) dry cutting with a band saw or circular saw to preserve alteration layers; (iii) subsequent analysis by micro X-ray diffraction (μ-XRD) and micro X-ray fluorescence (μ-XRF) for glass crystal structure and composition; and (iv) focused ion beam-scanning electron microscopy (FIB-SEM), followed by scanning transmission electron microscopy with energy dispersive spectroscopy (STEM-EDS) to analyze alteration layers.

Results and Discussion

The sample selected for investigation is composed of the local granitic gneiss and partly melted amphibolite. μ-XRD of exposed clear and dark glass regions suggest that they are predominantly amorphous, with a very minor crystalline component. From μ-XRF, Na, K, Al and Si represent the major elemental constituents of the clear glass, whereas the dark glass is enriched in Fe and Ca. SEM images show evidence of significant microbial activity on the glass surface. Preliminary investigations suggest that this microbial community is composed of bacteria, vitricolous lichen, fungal hyphae with associated mineral precipitates, and testate amoebae. To investigate alteration of clear and dark glass, in the presence and absence of microbial colonization, FIB cross-sections were analyzed by STEM-EDS. A FIB section was extracted from an area with a range of melting and solidification textures, including relict granitic solids and dendritic crystallization, and the clear glass melt area showed very little alteration. In contrast, a FIB-section taken from a bead-like area of clear glass, with evidence for microbial colonization on the surface, showed semi-circular alteration patterns, depleted in Na and K, on a micron-scale. Research has shown that colonizing microbes can alter the pH, resulting in localized glass dissolution. The formation of semi-circular alteration patterns on a micron scale has been observed as a result of microbial colonization and local dissolution in volcanic glass [13]. The dark glass region, near the clear glass, had areas with and without significant amounts of organic material on the surface. SEM images of the dark glass surface without organic material showed some pitting. The FIB section showed regions depleted and enriched in Fe and Ca, suggesting phase separation, but no semi-circular alteration patterns were evident.

Weaver, Jamie; Pearce, Carolyn; Vicenzi, Edward; Lam, Thomas; DePriest, Paula; Koestler, Robert; Varga, Tamas; Miller, Micah; Arey, Bruce; Conroy, Michele; McCloy, John; Sjoblom, Rolf; Schweiger, Michael; Peeler, David; Kruger, Albert.
"Investigating Alteration of Pre-Viking Hillfort Glasses From the Broborg Hillfort Site, Sweden."
Paper presented at Materials Science & Technology 2017 (MS&T 17), Pittsburgh, PA, United States. October 8, 2017 - October 12, 2017.

Koestler *et al.* defined the role of the surrounding environment on glass alteration in terms of three agents; physical, chemical and biological. Under natural conditions, these agents can occur independently or in combination to produce the alteration patterns observed [14]. Physical, chemical and biological factors are also expected to influence nuclear waste glass alteration [15]. For the Broborg glasses, the impact of each factor on glass alteration can be estimated based on knowledge of sample age, variations in climate, water activity and land use. SEM suggests that microorganisms interact with the glass surfaces, and may play a significant role in glass alteration, as they do in chemical weathering of the Earth's upper crust [16]. The microorganisms that are present provide information on the chemistry of the surrounding soil (water activity, nutrient concentrations, temperature and pH) and on cyclic weather patterns (freeze-thaw, wet-dry, and hot-cold conditions). A detailed study of the climate and environment around the Broborg site over the last ~1500 years is under way to understand this relationship. This will be combined with analysis of fresh glass and soil samples obtained from Broborg, to confirm preliminary findings and to test microbial alteration theories. Experiments are also being conducted to replicate the glassy phases for corrosion studies to validate current alteration test methods and models.

The glasses from Broborg fulfill several important prerequisites for good analogues for nuclear waste glass: a similar chemical compositional space, similar mechanisms of corrosion, and alteration in similar, known environmental conditions.

References

1. Goldschmidt, V.M., *The principles of distribution of chemical elements in minerals and rocks.* Journal of the Chemical Society (Resumed), 1937: p. 655-673.
2. Poinssot, C. and S. Gin, *Long-term Behavior Science: The cornerstone approach for reliably assessing the long-term performance of nuclear waste.* Journal of Nuclear Materials, 2012. **420**(1): p. 182-192.
3. Libourel, G., et al., *The use of natural and archeological analogues for understanding the long-term behavior of nuclear glasses.* Comptes Rendus Geoscience, 2011. **343**(2): p. 237-245.
4. Miller, W., et al., *Geological disposal of radioactive wastes and natural analogues*. Vol. 2. 2000: Elsevier.
5. Ewing, R., *Natural glasses: analogues for radioactive waste forms*, in *Scientific basis for nuclear waste management*. 1979, Springer. p. 57-68.
6. Sjöblom, R., H. Ecke, and E. Brännvall, *Vitrified forts as anthropogenic analogues for assessment of long-term stability of vitrified waste in natural environments.* International Journal of Sustainable Development and Planning, 2013. **8**(3): p. 380-399.
7. Sjöblom, R., et al. *Vitrified hillforts as anthropogenic analogues for nuclear waste glasses: project planning and initiation*. in *International Journal of Sustainable Development and Planning*. Vol. 11, No. 6 (2016) 897–906
8. Kresten, P. and B. Ambrosiani, *Swedish vitrified forts–a reconnaissance study.* Fornvännen, 1992. **87**: p. 1-17.
9. Friend, C.R., et al., *New field, analytical data and melting temperature determinations from three vitrified forts in Lochaber, Western Highlands, Scotland.* Journal of Archaeological Science: Reports, 2016. **10**: p. 237-252.
10. Kresten, P., C. Goedicke, and A. Manzano, *TL-dating of vitrified material.* Geochronometria, 2003. **22**: p. 9-14.
11. Kresten, P., L. Kero, and J. Chyssler, *Geology of the vitrified hill-fort Broborg in Uppland, Sweden.* GFF, 1993. **115**(1): p. 13-24.
12. Kresten, P., L. Larsson, and E. Hjärthner-Holdar, *Thermometry of ancient iron slags from Sweden, and of vitrified material from various hill-forts in western Europe.* Geoarchaeological Laboratory, UV Uppsala, 1996.
13. Hubert Staudigel, Harald Furnes, Nicola McLoughlin, Neil R Banerjee, Laurie B Connell, and Alexis Templeton, "3.5 billion years of glass bioalteration: Volcanic rocks as a basis for microbial life?," Earth-Science Reviews 89 (3), 156-176 (2008).
14. Koestler, R., et al., *Group report: how do external environmental factors accelerate change?*, in *Durability and change: the science, responsibility, and cost of sustaining cultural heritage*. 1994, John Wiley and Sons. p. 149-163.
15. Weaver, J.L., et al., *Ensuring Longevity: Ancient Glasses Help Predict Durability of Vitrified Nuclear Waste.* American Ceramics Society Bulletin, 2016. **95**(4): p. 18-23.
16. Konhauser, K., *Introduction to Geomicrobiology*, (Oxford, UK: Blackwell Publishing, 2007) 192.

Weaver, Jamie; Pearce, Carolyn; Vicenzi, Edward; Lam, Thomas; DePriest, Paula; Koestler, Robert; Varga, Tamas; Miller, Micah; Arey, Bruce; Conroy, Michele; McCloy, John; Sjoblom, Rolf; Schweiger, Michael; Peeler, David; Kruger, Albert.
"Investigating Alteration of Pre-Viking Hillfort Glasses From the Broborg Hillfort Site, Sweden."
Paper presented at Materials Science & Technology 2017 (MS&T 17), Pittsburgh, PA, United States. October 8, 2017 - October 12, 2017.

Pseudo-Exhaustive Verification of Rule Based Systems

D. Richard Kuhn[1], Dylan Yaga[1], Raghu N. Kacker[1], Yu Lei[2], Vincent Hu[1]

[1] National Institute of Standards and Technology
Gaithersburg, MD 20899, USA
{kuhn,dylan.yaga,raghu.kacker}@nist.gov

[2]Computer Science & Engineering
University of Texas at Arlington
Arlington, TX, USA
ylei@uta.edu

Abstract — **Rule-based systems are important in application domains such as artificial intelligence and business rule engines. When translated into an implementation, simple expressions in rules may map to a large body of code that requires testing. We show how rule-based systems may be tested efficiently, using combinatorial methods and a constraint solver in a test method that is *pseudo-exhaustive*, which we define as exhaustive testing of all combinations of variable values on which a decision is dependent. The method has been implemented in a tool that can be applied to testing and verification for a wide range of applications.**

Keywords — combinatorial testing; constraint solvers; formal methods; t-way testing; rule-based systems; test automation

I. Introduction

Rule-based systems have been important in a variety of application domains for many years. Some of the earliest artificial intelligence systems (AI) were designed to evaluate large rule sets, and this approach continues to be important for AI. In other domains, business rule engines automate complex enterprise resource planning (ERP) problems [1]. The terms used in rules may be expressed as Boolean (dichotomous) or relational conditions on inputs, values from databases, and environmental conditions such as time of day. Thus even a rule that contains only a few simple conditionals may invoke significant processing involved in computing the values used in the rule conditions. A rule-based system must work for any set of inputs, and can be implemented with a wide variety of rule engines. For example, JBoss, Oracle Policy Automation, OpenRules, Drools, IBM ODM, and many other tools exist to process rules supplied by users. But as with conventional software, exhaustive testing is nearly always intractable. This paper generalizes a practical method developed for testing access control systems [2], and introduces a tool that implements this method.

The approach to testing rule-based systems is *pseudo-exhaustive*, which we define as exhaustive testing of all combinations of variable values on which a decision is dependent. This approach is analogous to pseudo-exhaustive methods for testing combinational circuits [3], where the verification problem is reduced by exhaustively testing only the subset of inputs on which an output is dependent, or by partitioning the circuit and exhaustively testing each segment. The general concept of exhaustively testing subsets of variable values on which a decision is dependent is applied here to rule-based systems by transforming rule conditions to disjunctive normal form, then considering each term separately [2].

DOI reference number: 10.18293/SEKE2018-072

Testing a rule-based system requires showing that the rules as specified, P, are correctly implemented. The implementation P' must be shown to produce the same response as P for any combination of input values used in rules. That is, for input values $x_1,\ldots, x_n$, $P'(x_1,\ldots, x_n) = P(x_1,\ldots, x_n)$. Positive testing to show that a rule produces a specified result is easy: instantiate conditions to true for each antecedent associated with the result and verify that the system returns the designated result. Negative testing, showing that no combination of input values will produce the same result when it should not, is much more difficult. With n Boolean variables there are 2^n possible combinations of variables. For example, it would not be unusual to have 50 Boolean variables, resulting in $2^{50} \approx 10^{15}$ combinations, which would appear to make full negative testing intractable. In this paper, we show how combinatorial methods can be used to make this testing problem practical, given assumptions that apply to many or most rule-based systems.

II. Test Construction

We describe the derivation of complete test cases from rules converted to k-DNF structure (disjunctive normal form where no term contains more than k literals, and a *term* is a conjunction of one or more literals within the disjunction), using a constraint solver and a covering array generator. Two arrays are constructed for each possible rule consequent, such that every test in each array should produce the same result, with variations indicating an error. The method may be applied to rule systems with multiple outputs, where outputs are either discrete values or are defined by a predicate or expression with a Boolean result.

Rules are assumed to be given as expressions made up of variables with logical connectives in an antecedent, with a consequent given as a discrete value or simple predicate, structured as shown below where R_i are predicates evaluating the values of one or more variables, and $result_i$ is the result expected when conditions of R_i evaluate to true:

$$(R_1 \rightarrow result_1)\ (R_2 \rightarrow result_2)\ \ldots\ (R_m \rightarrow result_m)$$

Kuhn, David; Yaga, Dylan; Kacker, Raghu; Lei, Yu; Hu, Chung Tong.
"Pseudo-exhaustive Verification of Rule Based Systems."
Paper presented at 30th International Conference on Software Engineering and Knowledge Engineering (SEKE 2018), San Francisco, CA, United States. July 1, 2018 - July 3, 2018.

else → *default*

which is equivalent to:

$(R_1 \rightarrow result_1)\ (R_2 \rightarrow result_2)\ \ldots\ (R_m \rightarrow result_m)$

$(\sim R_1)\ (\sim R_2)\ldots\ (\sim R_m) \rightarrow$ *default*

Each R_i may include multiple variables, conditions, and logical connectives. It is required that the rule antecedents R_i are mutually exclusive, i.e., for any set of input variable values, only one antecedent will be matched. We believe this requirement is not overly restrictive, as in most applications it would be an error for matches of more than one rule. (It would be possible to use the constraint solver to check that rule antecedents are mutually exclusive, but this feature has not been implemented.)

Example 1: Suppose we have a rule set as shown below:

```
 if (a && (c && !d ||e))  R1;
else if (!a && b && !c)  R2;
else exit();
```

This code can be mapped to the following expression (note second line is "else", i.e., negation of predicates for R1 and R2):

$(a(c\bar{d}+e) \rightarrow R_1)\ (\bar{a}\ b\ \bar{c} \rightarrow R_2)$

$((\sim(a(c\bar{d}+e)))(\sim(\bar{a}\ b\ \bar{c})) \rightarrow exit)$

Literals can be conditions, such as age>18, or Boolean variables such as employee (yes, no), but the structure will be a series of expressions specifying subsets of conditions that produce each result, followed by a default rule when none of the attribute expressions have been instantiated to true.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
2	0	0	1	0	1	1	1	1	1	0	0	0	0	0	1
3	0	1	0	0	0	1	0	1	1	1	1	1	1	0	0
4	0	1	1	1	1	0	0	1	0	0	0	1	0	0	1
5	1	0	0	0	1	0	0	1	0	0	1	1	1	1	0
6	1	0	1	1	0	1	1	1	0	1	0	1	1	1	0
7	1	1	0	1	1	1	0	0	0	1	0	0	1	0	1
8	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1
9	1	0	0	0	0	1	1	0	1	0	0	1	1	1	1
10	0	1	1	1	0	1	1	0	1	0	1	1	0	1	0
11	0	1	0	0	1	0	1	0	0	0	1	0	1	1	1
12	1	0	1	1	1	0	0	0	1	0	1	0	1	0	0
13	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1
14	1	0	1	0	0	1	0	0	0	0	1	1	0	0	1
15	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0
16	0	1	1	0	1	0	0	0	1	1	0	1	1	1	1
17	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0
18	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
19	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0
20	1	1	1	1	0	0	1	0	0	1	1	1	0	1	1
21	1	0	0	0	1	1	0	0	1	0	0	0	0	1	0
22	0	1	1	1	0	1	1	1	0	1	0	0	1	1	1

Figure 1. 3-way covering array of 15 boolean parameters

To make testing tractable, we use combinatorial methods [2][4]. To see the advantages of a combinatorial approach, refer to Fig. 1, which shows a covering array of 15 boolean variables. A covering array is an $N \times k$ array of N rows and k variables. In every $N \times t$ subarray, each t-tuple occurs at least once. In software testing, each row of the covering array represents a test, with one column for each parameter that is varied in testing. For example, Fig. 1 shows a complete 3-way covering array that includes all 3-way combinations of binary values for 15 parameters in only 22 tests. The size of a t-way covering array of n variables with v values each is proportional to $v^t \log n$ [6][7]. For Example 1, with five attributes and two possible decisions for each attribute, there are $2^5 = 32$ possible rule instantiations. However, a covering array of all 3-way combinations contains only 12 rows. The number of variables for which all settings are guaranteed to be covered in a covering array is referred to as the *strength*; a 3-way array is of strength 3. We use covering arrays of variables from rules that have been converted to k-DNF form. For example, $abc + de$ contains two terms, one with three literals and one with two, so the expression is in 3-DNF form. The covering array does not contain all possible input configurations, but it will contain all k-way combinations of variable values. Where an expression is in k-DNF, any term containing k literals that is resolved to true will clearly result in the full expression being evaluated to true. For example, an access control rule in 2-DNF form could be: "if employee && US_citizen || auditor then *grant*". This rule contains one term of two attributes and one term of one attribute, so it is 2-DNF. Because a covering array of strength k contains every possible setting of all k-tuples and i-tuples for $i < k$, it contains every combination of values of any k literals.

As noted in the Introduction, we exhaustively test all combinations of values on which a decision is dependent. For the example above, the decision grant depends on either of two terms being true: employee && US_citizen or auditor. Any other setting of these three variables should result in deny. A truth table of all eight possible settings of these three variables would allow exhaustive testing of this set of rules. In general, exhaustive testing is intractable for nearly all applications, but note that at most two variables are required to produce a grant result. So if we test all 2-way combinations of settings of the input variables, we have achieved exhaustive testing of all combinations of variable values on which a decision is dependent, since no decision depends on more than two variables. (Later in the paper we show how this approach scales up to larger problems, and address the effectiveness for detecting errors when implemented rules contain more variables than are included in specified rules.)

Covering array generation tools, such as ACTS [4][6], make it possible to include constraints that prevent inclusion of variable combinations that meet criteria specified in a first order logic style syntax. For example, if we are testing applications that run on various combinations of operating systems and browsers, we may include a constraint such as 'OS = "Linux" => browser != "IE"'. Constraints are typically used in situations such as this, where certain combinations do not occur in practice or are physically impossible, and therefore should not be included in tests. Modern constraint solvers such as Choco [8] and Z3 [9] make it possible to process very complex constraint sets, converting logic expressions into combinations that are invalid and can be avoided in the final array.

Method: Let R = rule antecedents (left side of an implication rule such as p in $p \rightarrow q$) of one or more rules being tested in k-DNF, and T_i are terms (conjuncts of one or more variables or

Kuhn, David; Yaga, Dylan; Kacker, Raghu; Lei, Yu; Hu, Chung Tong.
"Pseudo-exhaustive Verification of Rule Based Systems."
Paper presented at 30th International Conference on Software Engineering and Knowledge Engineering (SEKE 2018), San Francisco, CA, United States. July 1, 2018 - July 3, 2018.

terms) in R. We designate the result/consequent of the rule being tested as (+), and any other possible result as (-). For the example included in Example 1, terms T_i of R_1 would be $ac\bar{d}$, and ae, and R_1 would be designated as (+) and R_2 or exit() designated as (-), for this test.

Positive testing: Generate a test set PTEST for which every test should produce a particular response. It must be shown that for all possible inputs, where some combination of k input values matches a (+) condition, a (+) result is returned. Construct test set PTEST = {P_{TESTi}} with one test for each term T_i of R as follows: $P_{TESTi} = T_i(\bigwedge_{j\neq i} \sim T_j)$

The construction ensures that each term in P is verified to independently produce the expected response for that rule. Negating each term T_j, $i \neq j$, prevents masking of a fault in the presence of other combinations that would return the same result. For example, if a rule condition is $ab + cd \rightarrow R_1$, inputs of 1100, 1101, 1110 could be used for testing $ab \rightarrow R_1$. However, input 1111 would not detect the fault if the system ignores variable a or b, because the condition cd would cause a result of R_1, and no other predicates in the rule would be evaluated. One such test is required for each term in a rule, so for m rules with an average of p terms each, the number of tests required is proportional to mp.

Negative testing: Generate a test set NTEST for which every test should produce a response other than the result designated by the rule being tested. It must be shown that for all possible inputs, where no combination of k input values matches a rule, an alternative result is returned.

> NTEST = covering array of strength k, for the set of variables in all rules, with constraints specified by $\sim R_i$.

Note that the structure of the rule evaluation makes it possible to use a covering array for NTEST, compressing a large number of test conditions into a few tests. Converted to k-DNF, each rule antecedent includes a sequence of conditions that are each sufficient to trigger the specified result. Because rule antecedents are mutually exclusive, masking of one combination by another can only occur for NTEST when a test produces a negative response, i.e., a response that is not a consequent of the rule instantiated in PTEST. In such a case, an error has been discovered, which can be repaired before running the test set again. Since NTEST is a covering array, the number of tests will be proportional to $v^k \log n$, for v values per variable (normally v=2 since most will be Boolean conditions), and n variables.

Rule antecedents are assumed to be mutually exclusive (to prevent masking as discussed above), but we allow for cases where multiple rules may have the same consequent (result). In such cases, rule antecedents are combined to produce the set of conjuncts used in generating PTEST and NTEST arrays. For example, if two rules are $R_1 \rightarrow Q_1$ and $R_2 \rightarrow Q_1$, then k-DNF terms for PTEST are produced from $(R_1 + R_2)$ and constraints for NTEST are given by $\sim(R_1 + R_2)$. For m rules with the same consequent (result), the number of tests is multiplied by the constant m.

Example 2: Table I gives a set of Boolean variables a through e, where each row defines values for the variables that determine an access control decision, either *grant* (+) or *deny* (-). Thus a covering array for the antecedent R of a rule in 3-DNF such as $(ac\bar{d} + \bar{a}b\bar{c} \rightarrow grant)$ is given in Table 1. The total number of 3-way combinations covered is the number of settings of three binary variables multiplied by the number of ways of choosing three variables from five, i.e., $2^3 \binom{5}{3} = 80$.

TABLE I. 3-WAY COVERING ARRAY

	a	b	c	d	e
1	0	0	0	0	0
2	0	0	1	1	1
3	0	1	0	1	0
4	0	1	1	0	1
5	1	0	0	1	1
6	1	0	1	0	0
7	1	1	0	0	1
8	1	1	1	1	0
9	1	1	0	0	0
10	0	0	1	1	0
11	0	0	0	0	1
12	1	1	1	1	1

TABLE II. 3-WAY COVERING ARRAY WITH CONSTRAINT ~R

	a	b	c	d	e
1	0	0	0	0	0
2	0	0	1	1	1
3	0	1	1	0	0
4	1	0	0	1	0
5	1	0	1	1	0
6	1	1	0	0	1
7	1	1	1	1	1
8	0	0	1	0	1
9	1	1	0	1	0
10	0	0	0	1	1
11	1	0	0	0	0
12	0	1	1	1	0
13	1	0	0	0	1
14	0	1	1	0	1

Table II shows a covering array for this set of variables generated using $\sim R$ as a constraint. That is, the two terms of the rule, $ac\bar{d}$ and $\bar{a}b\bar{c}$, have been excluded from the array, but all other 1-, 2-, and 3-way combinations can be found in the array. Because $ac\bar{d}$ and $\bar{a}b\bar{c}$ are the only conditions under which access should be granted, the array in Table II should result in a *deny* response from the system for every test. Collectively, tests include all 78 3-way settings of variables that will not instantiate the access control rule to *true*.

III. FAULT DETECTION PROPERTIES

Now consider the faults that this method can detect. Suppose that some combination of variables exists that produces a different response than required by the rule set P, for example because of errors in code that instantiates variable values. Tests contained in PTEST and NTEST will detect a large class of

Kuhn, David; Yaga, Dylan; Kacker, Raghu; Lei, Yu; Hu, Chung Tong.
"Pseudo-exhaustive Verification of Rule Based Systems."
Paper presented at 30th International Conference on Software Engineering and Knowledge Engineering (SEKE 2018), San Francisco, CA, United States. July 1, 2018 - July 3, 2018.

missing terms, added terms, or altered terms containing k or fewer variables. In this section we analyze faults that will be detected, and the underlying conditions in these faults. Table III illustrates the fault types and detection conditions for each.

TABLE III. EXAMPLE FAULTS AND DETECTION CONDITIONS.

	Term	C=correct term	F=faulty term	PTEST detect condition	NTEST detect condition	notes
1	**missing**	*abc*	--	*abc*	*none*	
2	**added**	--	*ab*	*none*	*ab*	
3		*abc*	$\bar{a}b$	*none*	$\bar{a}bc$, $\bar{a}b\bar{c}$	
4		*abc*	*ab*	*none*	$ab\bar{c}$	
5		*ab*	*abc*	--	--	*no fault*
6	**altered**	*abc*	$ab\bar{c}$	*abc*	$ab\bar{c}$	
7		*abc*	*ab*	*none*	$ab\bar{c}$	
8		*abc*	$\bar{a}b$	*abc*	$\bar{a}bc$, $\bar{a}b\bar{c}$	

k-DNF detection property: It is shown in [2] that collectively, tests from PTEST and NTEST will detect faults introduced by added, deleted, or altered terms with up to k variables. We can also show [2] that if more than k attributes are included in the altered term, some faults are still detected. Specifically, where a correct term has more than k variables and is not a subset of a faulty term, the fault will be detected. If a correct term is a subset of a faulty term in this case, some faults will be detected.

IV. SOFTWARE TOOL

The prototype research tool, Pseudo-Exhaustive Verifier (PEV), was developed in Java, and utilizes several open source external Java libraries. The software is packaged as a Java Archive (.jar) file which is directly executable as a Graphical User Interface (GUI), or can be run as a Command Line Interface (CLI) from a terminal.

The PEV software has been designed to accept rule sets comprised of Boolean variables, Boolean operators and relational expressions, implementing the algorithm described in Sect. II. The software parses the rule set, converts to Disjunctive Normal Form (DNF), inverts the DNF rule set, solves for positive conditions, and uses NIST's Automated Combinatorial Testing for Software (ACTS) tool [6] to compute a covering array for negative conditions.

Algorithm implementation: PEV utilizes several publicly available Java Archive libraries to generate test arrays. Transforming the input Boolean rule set to DNF is done using *jbool_expressions* [10], and the Choco constraint solver [11] is used for resolving relational statements.

Parsing: Parsing is a critical step of the PEV software, which occurs before any testing is performed. Since the software needs to accept input from the user, any input must be modified and sanitized prior to use, to ensure compatibility with the various APIs used, as well as to catch any syntactical problems prior to testing... The parser strips extraneous whitespace, and then normalizes Boolean operators `(&&, &, ||, |, !, ~)`, and attempts to match open and closing parenthesis. This sanitization ensures compatibility with the various APIs used throughout the software, and catches any syntactical problems prior to testing.

The software is not restricted to Boolean expressions, and has initial support for relational expressions (e.g., b < 3;). Note that a semicolon is used to identify a relational expression. During parsing, PEV will locate numeric relational expressions and replace them with temporary Boolean variables. After the replacement, the rule set is processed as normal. The relational values are solved at a later step and the results are recorded.

Once the initial input rule set is parsed, the software will convert it to Disjunctive Normal Form (DNF) to be tested. The user will be presented with a breakdown of the DNF rule set (split on the OR statements), each part of which is a positive condition that needs to be solved. Additionally, the user can set minimum and maximum values for any relational variable found in the rule set.

Solve for positive conditions: Each individual expression between OR operators is an expression that, once solved, will produce one positive condition. These expressions represent the only possible positive conditions for the original rule set – so it is possible to produce exhaustive positive conditions.

Consider Fig. 1, with the original input rule set:

```
emp & age>18; & (fa | emt | med) | b<3;
```

Converted to DNF, this is:

```
((age > 18; & emp & emt) | (age > 18; & emp
& fa) | (age > 18; & emp & med) | b < 3;)
```

Splitting on the OR operators, there are four individual expressions for the positive conditions (replacing relational expressions with temporary Boolean variables `tmp0 = age > 18; and tmp1 = b < 3;`):

- `tmp0 & emp & emt`
- `tmp0 & emp & fa`
- `tmp0 & emp & med`
- `tmp1`

To solve these expressions, any variable present is evaluated with the following rules, as shown in Table V:

- Non-negated variables evaluate to true
- Negated variables evaluate to false
- Variables not present evaluate to false

Solve for negative conditions: Depending on the complexity of the input rule set, it may not be feasible to produce exhaustive negative condition output combinations. By utilizing combinatorial test methods, it is possible to generate covering arrays of sufficient strength to have good test coverage. The method for producing negative conditions can be found by generating the full covering array for all the unique Boolean variables within the rule set, and using the DNF rule set as a constraint – which will remove the positive conditions from the resulting output.

Kuhn, David; Yaga, Dylan; Kacker, Raghu; Lei, Yu; Hu, Chung Tong.
"Pseudo-exhaustive Verification of Rule Based Systems."
Paper presented at 30th International Conference on Software Engineering and Knowledge Engineering (SEKE 2018), San Francisco, CA, United States. July 1, 2018 - July 3, 2018.

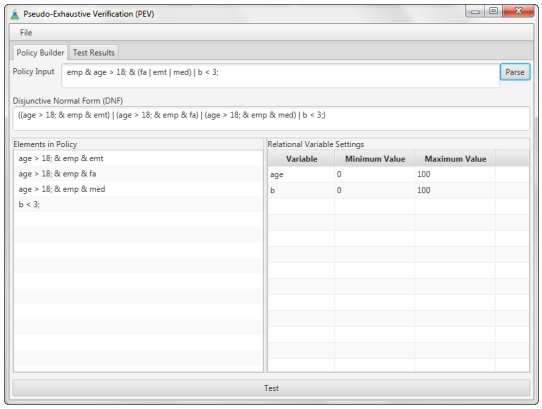


Figure 1. PEV software, after initial rule set parsed

TABLE V - SOLVED POSITIVE CONDITIONS

Expression	tmp0	tmp1	emp	emt	fa	med
tmp0 & emp & emt	1	0	1	1	0	0
tmp0 & emp & fa	1	0	1	0	1	0
tmp0 & emp & med	1	0	1	0	0	1
tmp1	0	1	0	0	0	0

A covering array for all negative conditions is computed as described in Sect. II. To perform this task, PEV creates an internal instance of the ACTS software, and passes a list of the unique Boolean variables from the rule set (including temporary Boolean replacements for relational expressions). The next step is to add the DNF rule set as a constraint to the system – so that the positive conditions are not included as negative results. Finally, the *k*-way combination is dynamically set after the *k*-DNF transform, which finds the conjunction with the largest combination of Boolean variables. In this example, the value of 3 is set (Table VI). PEV currently supports $k = 2..6$, because ACTS is used as the covering array generator, but there is no inherent limit to 6-way combinations and the method could support $k > 6$.

Solve for relational expressions using the Choco Expression Parser and the Choco Constraint Solver.

Relational Expression Formatting: The general format is:

```
Variable OPERATOR Integer_Value; Or
Integer_Value OPERATOR Variable;
```

Every relational expression must end with a semicolon (;), and two or more relational expressions in a row (without Boolean operators between them) will be replaced with one temporary Boolean variable during parsing. An example is shown in Table VII.

After being extracted and replaced by temporary Boolean variables, and the Positive/Negative conditions are found, an instance of Choco Expression Parser is created, and the relational expressions are passed as parameters. The minimum and maximum range for the expression to test against must be set – the PEV GUI includes a section which will allow the adjustment of every relational variable min and max values (default set to 0 to 100). These values can be adjusted prior to testing the rule set so that a customized range can be found. The solutions to the solved expressions are then placed into the results where appropriate.

TABLE IV. SOLVED NEGATIVE CONDITIONS

tmp0	tmp1	emp	emt	fa	med
1	0	1	0	0	0
1	0	0	1	1	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	0	1	0	0
0	0	1	0	1	1
1	0	0	0	1	0
0	0	1	1	0	1
1	0	0	1	0	1
0	0	0	0	1	0
1	0	0	0	0	1
1	0	0	1	0	0

TABLE V. INPUT POLICIES AND RESULTING PARSED RULE SET

Input Rule set	Parsed Rule set
a > 10; 20 < b; || n	tmp0 || n
a > 10; || 20 < b; || n	tmp0 || tmp1 || n

Results: Once testing completes, PEV displays usage metrics and parameters which will result in positive conditions, and the covering array for negative conditions. At this point, the results can be saved as a comma separated value (.csv) file.

V. V. TEST SET SIZE AND PRACTICAL IMPLICATIONS

The process scales easily to systems with a large number of variables and rules. Because the number of rows in a covering array grows only with $\log n$ for n variables at a given number of values, a large increase in the number of variables requires only a few additional tests.

The most significant limitation for this approach occurs where terms in rules contain a large number of values per variable. Because the number of rows of a covering array increases with v^k, for v variable values, if terms in the rules have more than 10 to 12 values, it may not be practical to generate covering arrays. However, a large number of tests is not a barrier, because the structure of the solution resolves the oracle problem by ensuring that every test in PTEST should produce a response of (+) and every test in NTEST should produce a response of (-). Consequently, tests can be fully automated, making it possible to execute a large test set.

VI. VI. RELATED WORK

This paper generalizes a method developed originally for testing attribute-based access control systems [2], which had been incorporated into the Access Control Policy Testing tool ACPT [12]. The generalized method and new tool, PEV, were developed to make the method useful in development and testing for a wider range of applications. Pseudo-exhaustive test methods for circuit testing have an extensive history of application [1]. While our method is not derived from these earlier approaches, it shares the basic notion of determining dependencies, partitioning according to these dependencies, and

Kuhn, David; Yaga, Dylan; Kacker, Raghu; Lei, Yu; Hu, Chung Tong.
"Pseudo-exhaustive Verification of Rule Based Systems."
Paper presented at 30th International Conference on Software Engineering and Knowledge Engineering (SEKE 2018), San Francisco, CA, United States. July 1, 2018 - July 3, 2018.

testing exhaustively the inputs on which an output is dependent. We have previously applied this notion to software testing in a more general form, using the observation that faults depend on a small number of inputs, by covering all 2-way to 6-way combinations of inputs [13]. This earlier work generated a test oracle using a model checker with a formal specification of a system, instantiated with inputs from a covering array.

Relatively little work has been published on testing specifically for rule-based systems. Dalal et al. [15] describe a case study of a rule-based system in an evaluation of model based testing, including the use of the combinatorial testing tool AETG. However, their testing considered only high level properties, such as whether updates correlated with the assignment of jobs during a working day. That is, no tests were generated from the rules. Rule based systems have also been used in a number of studies of test data generation [16][17], but used rules in generating tests for other software, rather than testing the rule-based systems themselves.

Among automated test generation systems, PEV falls into the class of tools with a specified test oracle, using the taxonomy of Barr et al. [14], because system rules serve as a specification of system behavior. Many such systems have been developed. The test oracles used in those systems were designed to answer the question "For a given set of inputs and initial state, what is the system output?", using a formal spec of some kind. Given such an oracle, test inputs must also be provided. Our method differs from these in that we address a narrower class of systems, but trade this limitation for complete coverage of inputs up to k-way combinations, providing testing that is pseudo-exhaustive, i.e., exhaustive for all subsets of inputs on which a rule result is dependent.

VII. Conclusions

Rule-based systems are used extensively in applications such as enterprise resource planning and machine learning [20]. If rules contain at most k Boolean variables per conjunction, for an expression in k-DNF, then a k-way covering array can test all possible settings of such terms. Thus for any possible combination of n inputs, only k ($k < n$) matter in determining the truth of the expression. In most applications, the number of conditions in conjunction will be small, even though the number of rules may be very high, possibly several hundred or even into thousands. The number of rows in a k-way covering array of Boolean variables is proportional to $2^k \log n$, and the ACTS covering array generator used in PEV produces arrays up to 6-way. Therefore PEV can efficiently process thousands of conditions or rules with up to six conditions per conjunction, sufficient for practical use.

The method described here was initially used in access control policy testing [2], and PEV has extended its applicability to a broader range of potential use. We are also considering methods to improve the efficiency of the PEV tool, including use of SAT solvers for generating covering arrays [18][19]. It may be possible to integrate the methods described in this paper with SAT-solver based covering array generation, to produce more compact arrays.

To make the tool more useful for practical application, features to allow import and export from common rule system formats, or decision table structures, may be helpful. We plan to investigate the possibilities depending on interest from users. We have received inquiries regarding compatibility with commercial tools, which could be considered for further development. Thus far, the major interest for this test method is for business rule systems, but it could be applied to traditional expert system applications as well.

Note: *Identification of products does not imply endorsement by NIST, nor that products identified are necessarily the best available for the purpose.*

References

[1] Lu, R., & Sadiq, S. A survey of comparative business process modeling approaches. In *Intl Conf on Business Information Systems* (pp. 82-94). Springer, 2007.

[2] Kuhn, D. R., Hu, V., Ferraiolo, D. F., Kacker, R. N., & Lei, Y. (2016, April). Pseudo-exhaustive testing of attribute based access control rules. In *Software Testing, Verification and Validation Workshops (ICSTW), 2016 IEEE Ninth International Conference on* (pp. 51-58).

[3] McCluskey, E. J. (1984). Verification Testing: A Pseudoexhaustive Test Technique. *Computers, IEEE Transactions on*, *100*(6), 541-546.

[4] Kuhn, D. R., Kacker, R. N., & Lei, Y. (2010). SP 800-142. Practical Combinatorial Testing, NIST, Gaithersburg, MD 20899

[5] ACTS Home Page, http:// csrc.nist.gov/acts/

[6] Y. Lei, R. Kacker, D.R. Kuhn, V. Okun, J. Lawrence, IPOG: A general strategy for t-way software testing. *14th intl conference on the engineering of computer-based systems*, 2007, pp 549–556

[7] D. M. Cohen, S. R. Dalal, M. L. Fredman, and G. C. Patton, "The AETG System: An Approach toTesting Based on Combinatorial Design," *IEEE Trans. Software Eng.*, 23(7):437-444,1997.

[8] Jussien, N., Rochart, G., & Lorca, X. (2008). Choco: an open source java constraint programming library. *Open-Source Software for Integer and Contraint Programming (OSSICP'08)* (pp. 1-10).

[9] De Moura, L., & Bjørner, N. (2008). Z3: An efficient SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems* (pp. 337-340). Springer Berlin Heidelberg.

[10] https://github.com/bpodgursky/jbool_expressions

[11] https://github.com/kaktus40/choco-exppar
http://www.choco-solver.org/

[12] ACPT Home Page, http://csrc.nist.gov/groups/SNS/acpt/access_control_policy_testing.html

[13] D. R. Kuhn, V. Okun, Pseudo-exhaustive Testing For Software, 30th NASA/IEEE Software Engineering Workshop, April 25-27, 2006

[14] Barr, E. T., Harman, M., McMinn, P., Shahbaz, M., & Yoo, S. (2015). The oracle problem in software testing: A survey. *IEEE transactions on software engineering*, *41*(5), 507-525.

[15] Dalal, S. R., Jain, A., Karunanithi, N., Leaton, J. M., Lott, C. M., Patton, G. C., & Horowitz, B. M. (1999, May). Model-based testing in practice. *21st Intl Conf on Software Eng.* (pp. 285-294). ACM.

[16] Deason, W. H., Brown, D. B., Chang, K. H., & Cross, J. H. (1991). A rule-based software test data generator. *IEEE transactions on Knowledge and Data Engineering*, *3*(1), 108-117.

[17] Edvardsson, J. A survey on automatic test data generation. *2nd Conference on Computer Science and Engineering* (pp. 21-28) 1999.

[18] Lopez-Escogido D, Torres-Jimenez J, Rodriguez-Tello E, Rangel-Valdez N. Strength two covering arrays construction using a sat representation. InMICAI 2008: Advances in Artificial Intelligence 2008 Oct 27 (pp. 44-53). Springer Berlin Heidelberg.

[19] Banbara M, Matsunaka H, Tamura N, Inoue K. Generating combinatorial test cases by efficient SAT encodings suitable for CDCL SAT solvers. InLogic for Programming, Artificial Intelligence, and Reasoning 2010 Oct 10 (pp. 112-126). Springer.

[20] Lee, C. C. (1991). A self-learning rule-based controller employing approximate reasoning and neural net concepts. *International Journal of Intelligent Systems*, *6*(1), 71-93.

Kuhn, David; Yaga, Dylan; Kacker, Raghu; Lei, Yu; Hu, Chung Tong.
"Pseudo-exhaustive Verification of Rule Based Systems."
Paper presented at 30th International Conference on Software Engineering and Knowledge Engineering (SEKE 2018), San Francisco, CA, United States. July 1, 2018 - July 3, 2018.

Investigation of Alteration of ^{6}Li Enriched Neutron Shielding Glass

Jamie L. Weaver, Danyal Turkoglu

National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD, 20899

INTRODUCTION

Silicate glass doped with ^{6}Li is a common slow neutron shielding material, and has been utilized as such in several neutron research facilities. ^{6}Li is a unique isotope for thermal neutron capture as it has a large thermal neutron capture cross section (≈ 941 b), and the compound nucleus primarily decays via ^{6}Li(n, α)^{3}H. This nuclear reaction produces a minor prompt-gamma ray branch (0.004 %, ≈ 37 mb), but otherwise meets the requirement that neutron shielding materials should produce minimal gamma-ray emissions [1]. Although ^{6}Li has a low natural isotopic abundance (7.5 %), it is commercially available at enrichments greater than 95 % in the United States.

Of the variety of ^{6}Li materials proposed for neutron shielding, a review of which can be found in Ref. [1], ^{6}Li enriched silicate glass is preferred as it provides a large specific neutron attenuation due to high ^{6}Li content, and can be fabricated into a variety of shapes and sizes with polished surfaces. Additionally, as the silicate glass structure can readily accept other elements (e.g., B and Al) the glass can be formulated to include elements that will improve its formability and durability.

The radiation durability of ^{6}Li silicate glass has been studied in detail, but the chemical durability has received less attention. The durability of the glass is important considering silicate glasses can form a hydrated, gel-like surface layer when exposed to a humid atmosphere [2]. This alteration layer can incorporate H and be depleted in alkaline elements relative to the unaltered glass. The result of these chemical reactions can be the slow depletion of ^{6}Li in the bulk glass with time, and/or a change in the distribution of ^{6}Li across the glass surface. Both outcomes could lead to a change in the glass's neutron shielding properties.

In this study at the National Institute of Standards and Technology (NIST) Center for Neutron Research (NCNR), the alteration of a ^{6}Li silicate glass designed for neutron shielding (NIST K2959 [1]) that had been exposed to an ambient environment for ≈ 3 years was investigated by a Neutron Depth Profiling (NDP) and Prompt Gamma-ray Activation Analysis (PGAA). NDP measures the concentration profile in the near surface for samples containing isotopes (e.g., ^{6}Li, ^{10}B) that undergo charged-particle (*i.e.*, *p, t,* or *α*) emission following neutron capture. The residual energy of the detected charged particles yields the energy lost within the sample before traveling in high vacuum to the detector. From stopping power curves and this residual energy, the depth at which the reaction took place is determined.

PGAA is a bulk technique for isotopic and elemental analysis based on gamma-ray spectroscopy of characteristic prompt gamma-ray emissions that immediately follow neutron capture. While PGAA is suitable for most naturally-occurring elements, it is unique in its ability to nondestructively quantify H content. NDP and PGAA provide complementary information in the study of the alteration of ^{6}Li glass.

DESCRIPTION OF THE ACTUAL WORK

The ^{6}Li glass was from a batch of glass produced for PGAA collimators and as a liner for the PGAA sample chamber. Since low-level H measurements by PGAA are a critical application, the ^{6}Li glass was intended to be a hydrogen-free material for absorbing scattered neutrons [3]. However, alteration of the glass surface, and subsequent increase in hydrogen, was suspected due to the surface's cloudy appearance. This was contrary to the clear appearance of the glass's surface when it was first cast. The glass sample analyzed had not been exposed to radiation prior to this analysis nor extensively handled after being cast.

Two indications of glass alteration are depletion of alkali from a glass surface, and an increase in the H content of the surface [2]. A glass surface can become depleted in alkali as glass network bound Si reacts with water. This reaction, which has been discussed in detail in Ref. [2], can lead to a restructuring of the glass network to accommodate the incorporation of H, and the release of glass network modifying ions, such as alkali, from the glass network.

The ^{6}Li enriched glass was tested for alteration by using NDP to measure the relative concentration ^{6}Li at and near the glass surface, and PGAA to measure the amount of H. To see how far the possible alteration had penetrated the glass, one side of a sample of the altered glass had (36.8 ± 0.3) μm of the surface removed by dry sanding with variably-decreasing grits (200 μm, 100 μm, 50 μm, 10 μm, and 2 μm grits) of silica-carbide paper. The amount of material removed was measured in triplicate at three points along the glass's surface with a digital micrometer (293 MDC-MX Lite, Mitutoyo). This sanding was necessary as the first NDP results (detailed below) showed that the alteration layer extended beyond the measurement ability of the NDP. Removing part of the alteration layer allowed for deeper measurement into the

Weaver, Jamie; Turkoglu, Danyal.
"Investigation of Alteration of 6 Li Enriched Neutron Shielding Glass."
Paper presented at the 2017 American Nuclear Society Winter Meeting, Washington, D.C., United States. October 29, 2017 - November 2, 2017.

glass by NDP. Dry sonication, in an inverter position (sanded surface facing down to allow loosened materials to fall away from the surface), followed by a cleaning with compressed and oil-filtered nitrogen gas was used to remove sanding debris from the glass surface. This cleaning process was repeated twice to ensure the surface was free of debris. No solid or liquid lubricants were used in the surface preparation process as there was a concern that the chemicals may remove Li or part of the alteration layer from the glass. A subsequent section of the glass sample was broken to expose a fresh, unaltered glass surface. As compared to the altered surface and the sanded surface, the broken surface should have relatively little water on its surface. The altered surface, sanded surface, and freshly-broken surface were analyzed to determine the extent of alteration of the glass.

The 20 MW research reactor in the NCNR with a liquid hydrogen cold source at 20 K provides cold neutrons that are transported to experimental stations in the guide hall via super-mirror guides. NDP spectra were acquired at the Neutron Guide 5, Cold Neutron Depth Profiling station. A circular aperture made of 0.5 mm thick Teflon® fluorinated ethylene propylene (FEP) with a 10.0 mm diameter opening was mounted to an Al disk with a large (> 30 mm) hole in its center. A minimum of two spots on each sample surface were analyzed to determine if alteration was consistent across the glass's surface. Each sample was irradiated at a near constant fluence rate of cold neutrons; any variations were corrected via a neutron monitor during data processing. All experiments were conducted under vacuum and at room temperature. NDP spectrum were collected for ≈ 4 h per spot. Both ^{6}Li nuclear reaction products, α and triton (t) particles, were detected using a circular transmission-type silicon surface-barrier detector that was positioned ≈ 120 mm from the sample surface. Each spectrum was corrected for dead time (≈ 1.4 %) and background noise. A detailed description of the NDP setup and data processing steps can be found in ref. [4], however, the neutron guide at which the instrument was located has changed since the publication of the cited article. The triton interaction with the glass was modeled in SRIM (2013), assuming a pristine glass density of 2.42 g/cm^2 [5]. ^{6}Li concentrations were calculated in reference to the known concentration of ^{10}B in a B-implanted concentration standard (in-house), according to Eq. 1:

$$[a] = [b]\frac{\sigma_{0,b}}{\sigma_{0,a}}\frac{\rho_b}{\rho_a} \qquad (1)$$

where $[a]$ and $[b]$ are the concentrations (atoms/cm^2) of isotopes a and b being measured in the sample and standard, respectively, σ_0 is the thermal neutron cross-section for the charged-particle emission, and ρ is the count rate for the isotope being measured.

NDP results of the altered, sanded, and freshly broken surfaces are shown in Figure 1. The altered glass shows depletion of ^{6}Li at the surface of the glass. The depletion depth for this sample is greater than the detection depth of the NDP. The sanded glass profile shows a slightly lower concentration of ^{6}Li at its surface than the freshly-broken glass sample. The slight depletion in ^{6}Li between the sanded and freshly-broken glass suggests that the entire depleted layer of the glass was not removed by sanding, and that the alteration layer is > 30 µm thick.

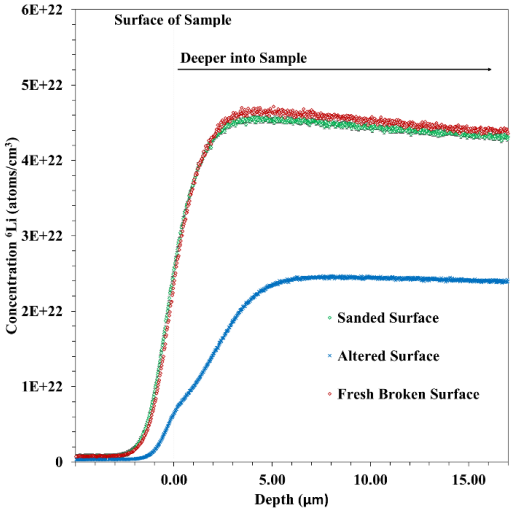


Fig. 1. NDP profiles (t profile only) of ^{6}Li concentration vs. depth of the altered glass, sanded glass, and fresh broken glass surfaces. The altered glass shows depletion of ^{6}Li at its surface, indicating glass alteration. Error bars relative to the concentration of ^{6}Li in each sample are reported to two (2) standard deviations.

The x-axis shown in Figure 1 was calculated assuming a glass density of 2.42 g/cm^3 for the sanded and freshly broken glasses, and 1.50 g/cm^3 for the altered glass. The density for the altered glass is based on densities reported for alteration layers on glasses of similar chemical compositions to the one in this study [6, 7]. Densities for the reported alteration layers are an average of total surfaces, and do not account for gradient changes in alteration layer densities that can occur over their profile (i.e. from surface of the alteration layer to the alteration layer/unaltered glass interface). The sanded glass may consist of, and have been measured in NDP, two or more layers of material; one or more that are altered, and one that is unaltered. The unaltered layer will most likely have a density of 2.42 g/cm^3, and the altered densities will most likely be slightly lower than that of the unaltered glass. Further analyses, *e.g.*, SEM, are needed to investigate the possible presence of these layers. Additionally, there is a slight decrease in the concentration of ^{6}Li as a function of depth in the sanded and fresh broken sample spectra. This is either due to a slight density or elemental concentration

Weaver, Jamie; Turkoglu, Danyal.
"Investigation of Alteration of 6 Li Enriched Neutron Shielding Glass."
Paper presented at the 2017 American Nuclear Society Winter Meeting, Washington, D.C., United States. October 29, 2017 - November 2, 2017.

variability within the glass. These kinds of variability can occur in a glass during its manufacturing, and is not unexpected.

Elemental analysis of the ^{6}Li glass was performed at the cold-neutron PGAA instrument located at Neutron Guide D in the NCNR. The ^{6}Li glass was heat-sealed within a Teflon FEP bag, and then suspended from aluminum wires spanning the sample holder frame. The sanded and altered portions of glass were measured separately to determine nominal changes in elemental concentration between the altered and sanded surfaces. The samples were measured for 3.0 h and 18.0 h, respectively, both with 1.4 % dead time. A blank Teflon FEP bag was measured for 3.9 h with 0.3 % dead time. Full-energy detection efficiency (ε) was calibrated from 50 keV to 11 MeV with measurements of decay gamma rays from ^{133}Ba and ^{152}Eu sources and prompt gamma rays emitted by N, Cl and Ti in urea, NaCl and Ti foil samples, respectively. Peak fitting and detection efficiency calibration were performed with Hypermet PC [8].

The atom ratios of element x to element y were determined using the relative approach shown in Eq. 2, where ρ_γ is the peak count rate, $\varepsilon(E_\gamma)$ is the detection efficiency and σ_γ is the partial gamma-ray production cross section [9]. The elemental cross section values were taken from the Evaluated Gamma-ray Activation File (EGAF) for H, Al, and Si [10]. Updated isotopic cross sections for ^{6}Li and ^{7}Li were used to determine the enrichment of ^{6}Li to be (96 ± 1) atom %. Atom ratios of Li/Si and Al/Si were determined using Eq. 2 [11].

$$\frac{n_x}{n_y} = \frac{(\rho_{\gamma_1,x} - \rho_{\gamma_1,blank})/\varepsilon(E_{\gamma_1})}{(\rho_{\gamma_2,y} - \rho_{\gamma_2,blank})/\varepsilon(E_{\gamma_2})} \frac{\sigma_{\gamma_2,y}}{\sigma_{\gamma_1,x}} \qquad (2)$$

The atom fractions, expressed as oxides, of Al, Si, and Li in the glass, were determined. Table I shows the nominal glass composition measured on the altered and sanded sides of the glass.

Table I: Comparison of the measured glass compositions of the altered and sanded sides of the ^{6}Li glass to the nominal formula as determined by PGAA.

Compound	Atom fraction (%)		
	Nominal [1]	Altered	Sanded
Li_2O	37	37.1 ± 0.8	37.2 ± 0.8
SiO_2	59	58.9 ± 0.8	58.8 ± 0.8
Al_2O_3	4	3.9 ± 0.1	4.0 ± 0.1

Gamma ray energies in keV for peaks used for analysis:
^{6}Li: 6769.5, 7246.7
^{7}Li: 2032.3
Si: 1273.3, 2092.9, 3539.0, 4933.9, 7199.2
Al: 1778.9, 3465.1, 4133.4, 4259.5, 4733.8

The background-corrected H count rates for the altered and sanded sides were (1.381 ± 0.014) counts s^{-1} and (0.186 ± 0.005) counts s^{-1}, respectively. The background was subtracted by using the peak ratio of H to F measured in the blank. As shown in Figure 2, the thermal-equivalent neutron flux decreases by more than an order of magnitude within 200 µm depth. Thus, the PGAA measurements of the ^{6}Li glass were weighted toward measuring the composition within the surface (i.e., < 1 mm depth). Quantifying the H content in the assumed alteration layer required modeling with MCNP6 to correct for the neutron self-shielding by the glass.

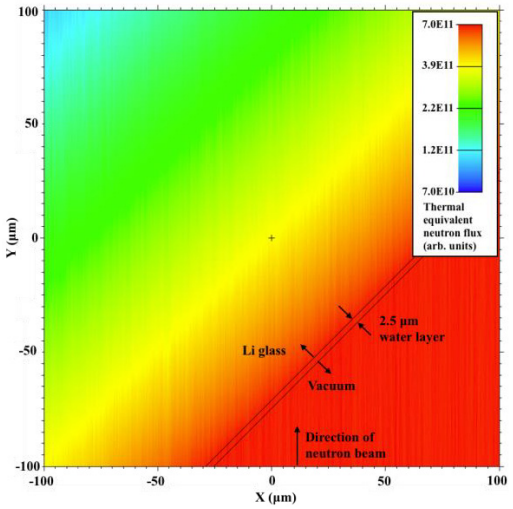


Fig. 2. MCNP6 mesh tally of the neutron flux weighted by a 1/v cross section (^{197}Au) for the ^{6}Li glass in the PGAA neutron beam.

The H content was estimated as the water-equivalent thickness on the surface of the ^{6}Li glass using Eq. 3. Here, the left side of the equation represents the ratio of efficiency- and background-corrected count rates for H and Si. The right side of the equation was solved with the MCNP6 [12] model with the following steps: 1) The H and Si reaction rates per unit volume R were estimated with neutron flux tallies for the water cell and the ^{6}Li glass cell, respectively. 2) The reaction rates were multiplied by the volumes, where A is the area and t_w and t_g are the thickness of the water and glass respectively, and multiplied by the probability for emission of a particular gamma-ray P_γ (which is obtained from $\sigma_\gamma = \sigma_0 \theta P_\gamma$, where θ is the natural isotopic abundance). 3) Since the reaction rates for H and Si were constant when various water layer thicknesses up to 20 µm were modeled, the t_w values were iterated until Eq. 3 was satisfied. The resulting water-equivalent thicknesses for the altered and sanded sides of the ^{6}Li glass were estimated with this approach to be 150 nm and 21 nm, respectively.

$$\frac{\rho_H/\varepsilon(E_{\gamma,H})}{\rho_{Si}/\varepsilon(E_{\gamma,Si})} = \frac{R_{H,w}\, A\, t_w}{R_{Si,g}\, A\, t_g} \frac{P_{\gamma,H}}{P_{\gamma,Si}} \qquad (3)$$

Weaver, Jamie; Turkoglu, Danyal.
"Investigation of Alteration of 6 Li Enriched Neutron Shielding Glass."
Paper presented at the 2017 American Nuclear Society Winter Meeting, Washington, D.C., United States. October 29, 2017 - November 2, 2017.

The PGAA results show that the surface of the ^{6}Li glass included H. Combined with the NDP results, which show a depletion of ^{6}Li at the surface of the glass relative to its bulk, the opaque nature of the glass, and the fact that much of the H was removed by sanding the surface of the glass, it is hypothesized that the surface of the ^{6}Li glass has been chemically altered from its original state. The most likely source of alteration is environmental exposure, predominately exposure to a humid atmosphere.

From the NDP results of the sanding experiment, it is estimated that the alteration layer thickness is > 30 μm. However, the MCNP6 modeling suggests a water-equivalent thickness near 150 nm. The MCNP6 model does not account for the fact that the H is most likely not a simple layer, but, in accordance with glass alteration theory [2], is probable incorporated into a Si rich alteration layer.

The amounts of ^{6}Li, Si, and Al nominally present in the glass is not statistically different from the amounts measured in the altered glass. This suggests that the introduction of H to the glass surface has not led to a release in ^{6}Li or other glass forming elements from the glass, only a rearrangement of the elements in the near-surface of the glass to accommodate H. Identifying the possible location of the ^{6}Li that was originally at the glass surface requires further research.

SUMMARY

Preliminary analysis by NDP and PGAA of a ^{6}Li glass used for neutron shielding was completed to investigate the alteration of the glass's surface. PGAA results show an increase in H in the altered sample relative to the sanded sample, although native glass forming element concentrations were determined to be the same (within error). NDP measurements indicate a depletion in ^{6}Li in the altered glass surface, and an alteration depth > 30 μm. These results confirm the hypothesis that the shielding glass is altered, and that the ^{6}Li concentration at the surface of the glass is less than expected. This leads to concerns in the efficacy of ^{6}Li silicate glasses as a neutron shielding material for PGAA. The increased H content in the glass may increase the overall H background signal for the instrument. Other applications for the glass, such as apertures or transmission application, may not be effected by the presence of a hydrated alteration layer. Further analyses are currently underway to characterize to determine the thickness, density, and morphology of the alteration layer.

ACKNOWLEDGMENT

The authors would like to thank Dr. Greg Downing and Dr. Heather Chen-Mayer for their mentorships and support in implementing this study.

DISCLAIMER

Trade names and commercial products are identified in this paper to specify the experimental procedures in adequate detail. This identification does not imply recommendation or endorsement by the authors or by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose. Contributions of the National Institute of Standards and Technology are not subject to copyright.

REFERENCES

1. C. Stone *et al.*, "^{6}Li-doped silicate glass for thermal neutron shielding," *Nucl. Inst. Meth. Phys. Res. A*, **349**, *2-3*, 515 (1994).
2. C. Cailleteau *et al*, "Insight into silicate-glass corrosion mechanisms," *Nat. Mater.*, **7**, *12*, 978, (2008).
3. D. Turkoglu *et al.,* "A ^{3}He beam stop for minimizing gamma-ray and fast-neutron background." *J. Radioanal. Nucl. Chem.*, **311**, *2*, 1243 (2016).
4. R. G. Downing, "Enhanced reaction rates in NDP analysis with neutron scattering," *Rev. Sci. Instrum.*, **85**, *4*, 045109 (2014).
5. J. Ziegler, "SRIM & TRIM", (2013).
6. J. T. Reiser *et al.*, "The use of positrons to survey alteration layers on synthetic nuclear waste glasses" *J. Nuc. Mtrl.*, **490,** *75*, (2017).
7. C. L. Trivelpiece *et al.*, "Glass Surface Layer Density by Neutron Depth Profiling", *Int. J. Apl. Gls. Sci.* **3**, 2, 137 (2012).
8. B. Fazekas *et al.*, "Introducing HYPERMET-PC for automatic analysis of complex gamma-ray spectra," *J. Radioanal. Nucl. Chem.*, **215,** *2*, 271-277 (1997).
9. G. Molnár *et al.*, "Prompt-gamma activation analysis using the k 0 approach," *J. Radioanal. Nucl. Chem.*, **234,** *1-2*, 21 (1998).
10. H. Choi *et al.*, "Database of prompt gamma rays from slow neutron capture for elemental analysis." (International Atomic Energy Agency) (2017).
11. R. Firestone and Z. Revay, "Thermal neutron radiative cross sections for ^{6,7}Li, ^{9}Be, ^{10,11}B, ^{12,13}C, and ^{14,15}N, *Phys. Rev. C*, **93**, *5*, 054306 (2016).
12. *MCNP® 6.1.1*, Los Alamos National Laboratory report LA–UR–14–24680, June 24, 2014

A Layered Graphical Model for Cloud Forensic and Mission Attack Impact Analysis

Changwei Liu[1], Anoop Singhal[2], and Duminda Wijesekera[1,2]

[1]Department of Computer Science, George Mason University, Fairfax VA 22030 USA

[2]National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg MD 20899 USA

[1]{*cliu6,dwijesek*}*@gmu.edu*

[2]*anoop.singhal@nist.gov*

Abstract

Cyber-attacks on the system supporting an enterprise's mission can impact achieving its objectives. We describe a layered graphical model as an extension of a forensic investigation in order to quantify mission impacts. Our model has three layers: the upper layer models operational tasks that constitute the mission and their inter-dependencies. The middle layer reconstructs attack scenarios from available evidence to reconstruct their inter-relationships. In cases where not all evidence is available, the lower level reconstructs potentially missing attack steps. Using the three levels of graphs constructed in these steps, we present a method to compute the impacts of attack activities on missions. We use NIST's National Vulnerability Database's (NVD)-Common Vulnerability Scoring System (CVSS) scores or forensic investigators' estimates in our impact computation. We present a case study to show the utility of our model.

Keywords: Mission attack impact, Enterprise infrastructure, Cloud forensic analysis, Layered-graphical model

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

1 Introduction

Organizational missions are used to abstract activities envisioned by organizations, usually defined at the high-level as a collection of business processes. Cyber-attacks on an enterprise infrastructure that support such missions can impact these missions. Concurrently, a growing number of organizational business processes and services are now hosted on cloud operators' data centers. Given that most networked infrastructures including cloud services are supported by hardware and software assets, any attack that impacts these assets could impact the missions they support. Therefore, analyzing and quantifying the mission impacts of cyber-attacks is of importance to infrastructure system planners in migrating security threats and improving mission resilience, which is the primary objective in this paper.

NIST's NVD-CVSS provides impact estimates of exploitable vulnerabilities on IT systems [2]. Other publications use NIST's NVD-CVSS to predict the impacts of multi-step attacks on assets by considering constructing all possible attack paths [1, 5, 6]. However, evaluating all paths is infeasible for forensic analysis to assess damages due to the large number of paths and vulnerabilities. Additionally, quoted publications only consider the vulnerabilities reported publicly, not including zero-day attacks.

Because post-attack artifacts obtained during forensic investigations provide information that can be used to analyze attacks, we create a layered graphical model that uses this information to quantify the attacks' impacts on missions. Our model has three layers. The upper layer models operational tasks that constitute the mission and their inter-dependencies, where a mission is modeled as a collection of choreographed tasks. The middle layer collects evidence from intrusion detection system (IDS) tools and event logs to reconstruct attack scenarios. In cases where not all evidence is available, the lower layer reconstructs potentially missing attack steps using system calls that were executed to fulfill the mission. Finally, the two mapping algorithms integrate the information obtained from the three layers to ascertain how the mission execution was supported midst attack activities. Using the layers of graph-like dependency information, our model provides a method to compute impacts of attacks on missions using the NIST NVD-CVSS scores or forensic investigators' estimates on attacks toward the underlying software or hardware. We present a case study to show the utility of our model, and how it can be used to migrate attack risks in a networked infrastructure. To the best of our knowledge, there isn't an integrated forensic analysis framework that quantifies the mission impacts of multi-step attacks in a complex enterprise infrastructure which uses cloud-based services.

The rest of the paper is organized as follows. Section 2 discusses background and related work. Section 3 presents the three-layered graphical model. Section 4 uses a case study to show how the model computes mission impacts of attacks in a cloud environment, and discusses how the impact analysis can be used to enhance a networked infrastructure. Section 5 gives the conclusion.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

2 Background and Related Work

We present the background and related work in this section.

2.1 Cloud Forensics

NIST defined Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) as deployment models [3]. SaaS allows clients to use providers' applications running on a cloud infrastructure. PaaS allows clients to deploy on the cloud client applications using programming languages, libraries, services and tools supported by the provider. IaaS provides clients with the capability of provisioning processing, storage, networks and other fundamental computing resources, so that the clients can deploy and run software including operating systems and applications.

Digital forensic investigators seek attack evidence from computers and networks. According to Ruan et al., cloud forensics is a subset of network forensics that follows the main phases of network forensics with techniques tailored to cloud computing environments [4]. For example, data acquisition is different in SaaS and IaaS, because the investigator will have to solely depend on cloud service providers in SaaS. Investigators can acquire the virtual machine images from IaaS customers.

2.2 Related Work

Attackers tend to use multi-step, multi-stage attacks to impact important services protected using complex mechanisms. Researchers have proposed and designed models to estimate the mission impacts of such attacks by considering all known vulnerabilities. Sun et al. proposed using a multi-layered impact evaluation model to estimate the mission impacts [10]. In this model, a lower vulnerability layer is proposed to map to an asset layer, and then to a service layer, which finally maps to the mission layer, so that the mission impacts can be calculated by using vulnerabilities' CVSS scores and the relationships between missions to the lower level vulnerabilities. However, this model does not provide a method to construct the attack paths. Another group of researchers, Sun et al., combined mission dependency graphs with attack graphs generated by an attack graph generation tool, MulVAL [11], to estimate the attack mission impacts in the clouds [6]. Noel at el. designed a cyber-mission impact assessment framework by leveraging Business Process Modeling Notation (BPMN) and their attack graph generation tool named Topological Vulnerability Analysis (TVA) [12] that combines an exploit knowledge base and a remote network scanner, analyzing all potential attack paths leading to attack goals to evaluate potential mission impacts [5, 6]. However, these approaches use vulnerabilities collected from the bug-report community such as NIST's NVD to assess the impacts of attacks. These do not scale to large infrastructures or zero-day attacks.

Forensics researchers have reasoned on post-attack evidence using correlation rules to reconstruct

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

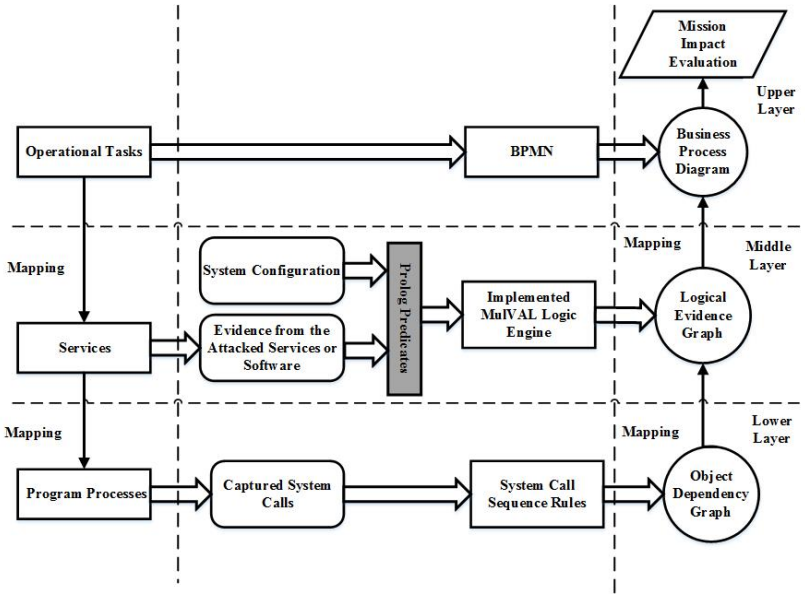


Figure 1: The three-layered graph model for mission impact evaluation

the attack scenarios. The objective of this work has been to reconstruct criminal or unauthorized actions shown to be disruptive to missions [7, 13]. To reconstruct attack scenarios that have legal standing, we integrated a Prolog logic tool, MulVAL, with two databases, including a vulnerability database and an anti-forensic database, to ascertain the admissibility of evidence and explain missing evidence due to attackers' using anti-forensics [9]. We also expanded this work by using system calls to reconstruct the missing attack steps due to missing evidence in the higher application level, and using Bayesian Network to estimate the experts' belief on the reconstructed attack scenarios [14]. However, no work exists to estimate the mission impacts of attacks launched toward an enterprise's infrastructure.

3 Our Three-layered Graphical Model

Figure 1 shows our model. The lower layers reconstruct attack paths so that the attacks can be mapped to tasks and missions in the upper layer for mission impact computation.

3.1 The Upper Layer

The upper layer models tasks and missions as business processes. We model business processes using a Business Process Diagram (BPD) using the Business Process Modeling Notation (BPMN). We use tasks, events, sequencing, exclusive choice, parallel gateways, message flows and pools [16] to constuct BPDs formalized in Definition 1 (Originally defined in [16]).

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

Definition 1 *Business Process Diagram-BPD:*

Any quintuple $(Pool, T, E, C, OP)$ *satisfying the following conditions are said to be a BPD.*

- *Let* T *be a set of tasks and* E *be a set of events.*
- *Let* E^{send}, E^{rec} *be two disjoint subsets of* E *and* $e_{start}, e_{end} \in E$ *be two events we refer to as start and end events. Hence,* $E = e_{start} \cup e_{end} \cup E^{send} \cup E^{rec}$ *are respectively called start, stop, message sending, message receiving events.*
- $T^{com} = \{(t, e_{send}, c), (t, e_{rec}, c) : t \in T, e_{send} \in E^{send}, e_{rec} \in E^{rec}, c \in C\}$ *are said to be the set of communicating tasks.*
- $OP = \{F, M, XOR, ;\}$ *called parallel fork, parallel merge, exclusive choice and sequencing.*
- C *is a set of channels.*

Business processlets and a Business Process are defined as follows.

- *Any* $t \in T \cup T^{com}$ *is said to be a processlet.*
- *If* P *and* Q *are processlets, then so are* $P; Q, F(P, Q)M$ *and* $P(XOR)Q$.
- *If* $P \in T \setminus T^{com}$ *and* e_{start}, e_{end} *are start and end events, then* $e_{start}; P; e_{end}$ *is said to be a business process.*
- *Pools are business processes with the constrains: for two pools* P_1 *and* P_2*, there is a task* (t_1, e_{send}, c) *in* P_1 *if and only if there will be another task* (t_2, e_{rec}, c) *in* P_2*, so that the message can be passed through channel* c.

3.2 The Middle Layer

The middle layer constructs potential attacks from evidence available from system logs and IDS alerts. Our objective is to map these attack scenarios to missions modeled as BPDs. Because we only consider attack scenarios substantiated using available evidence, the created attack paths do not include all possible attack paths and vulnerabilities. However, sometimes, not all evidence may be available in IDS logs, which will be addressed in the lower layer.

We reconstruct attack scenarios by using a forensic analysis tool created in our previous work [7]. This tool uses rules to create directed graphs of available items of evidence and correlate them together as witnesses of attacks. Because rules are used to create these graphs, they are called logical evidence graphs (LEGs) formalized in Definition 2, originally defined in [7]).

Definition 2 *Logical Evidence Graph-LEG:*

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

A $LEG = (N_r, N_f, N_c, E, L, G)$ *is said to be a logical evidence graph (LEG), where* N_f, N_r *and* N_c *are three sets of disjoint nodes called fact, rule, and consequence fact nodes respectively.* $E \subseteq N_f \cup N_c \times N_r \cup (N_r \times N_c$, *L is the mapping from a node to its labels and* $G \subseteq N_c$ *are observed attack events. Every rule node has a consequence fact node as its single child and one or more fact or consequence fact nodes from prior attack steps as its parents. The labels of nodes consist of instantiations of rules or sets of predicates specified as follows:*

1. *A node in* N_f *is an instantiation of predicates that codify system states including access privileges, network topology consisting of interconnectivity information, or known vulnerabilities associated with host computers in the system. We use the following predicates:*

   (a) *hasAccount(_principal,_host,_account), canAccessFile(_host,_user,_access,_path) etc. to model access privileges.*

   (b) *attackerLocated(_host) and hacl(_src,_dst,_prot,_port) to model network topology, namely, the attacker's location and network reachability information.*

   (c) *vulExists(_host,_vulID,_program) and vulProperty(_vulID,_range,_consequence) to model vulnerabilities exhibited by nodes.*

2. *A node in* N_c *represents a predicate that codifies the post attack state as the consequence of an attack step. We use predicates execCode(_host,_user) and netAccess(_machine,_protocol,_port) to model the attacker's capability after an attack step. Valid instantiations of these predicates after an attack will update valid instantiations of the predicates listed in (1).*

3. *A node in* N_r *consists of a single rule in the form* $p \leftarrow p_1 \wedge p_2, \ldots, \wedge p_n$ *with p as the child node of* N_r *is an instantiation of predicates from* N_c. *All* p_i *for* $i \in \{1 \ldots n\}$ *as the parent nodes of* N_r *are the collection of all predicate instantiations of* N_f *from the current step and* N_c *from the prior attack step.*

Table 1: Dependencies arising out of systems calls

Dependency	**Event Description**	**Unix System Calls**
$process \rightarrow file$	Process modifies file	write, pwrite64, rename, mkdir, linkat, link, symlinkat, etc
$file \rightarrow process$	Process reads file	stat64, lstat6e, fsat64, open, read, pread64, execve, etc.
$process \leftrightarrow file$	Process uses/modifies file	open, rename, mount, mmap2, mprotect etc.
$process1 \rightarrow process2$	Process1 creates/terminates Process2	vfork, fork, kill, etc.
$process \rightarrow socket$	process writes socket	write, pwrite64, etc.
$socket \rightarrow process$	process checks/reads socket	fstat64, read, pread64, etc.
$process \leftrightarrow socket$	Process reads/writes/checks socket	mount, connect, accept, bind, sendto, send, sendmsg, etc.
socket $\leftrightarrow$ socket	process reads/writes socket	connect, accept, sendto, sendmsg, recvfrom, recvmsg

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

3.3 The Lower Layer

The lower layer uses instances of interactions between services and the execution environment to obtain evidence unavailable from systems logs and IDS alerts. We obtain interaction instances from systems call logs. This usage is based on our postulate of missing evidence due to the attackers' using anti-forensic techniques, limitation of forensic tools, or zero-day attacks. Because there are many system calls, we use those chosen in [8], listed in the right-hand column of Table 1. Our abstraction of them appear in the left-hand column of Table 1. A process making system calls creates dependencies between itself and other processes, files, or sockets for network connection. We model these dependencies as graphs that we call object dependency graphs (ODGs), formalized in Definition 3.

Definition 3 *Object Dependency Graph-ODG:*

The reflexive transitive closure of $\rightarrow$ *defined in Table 1 is an object dependency graph. We use the notation ODG=*(V_O, V_E, E) *to represent an object dependency graph, where* V_O *is the set of vertexes that are composed of objects including Processes P, Files F or Sockets S;* V_E *is the set of textual event descriptions listed in the middle column; and E is the set of dependency edges listed in the left-hand column of Table 1.*

3.4 The Mapping between the Three Layers

The left-hand and right-hand columns in Figure 1 show the system resource mapping and graph mapping of our model. We use the resource mapping obtained from the infrastructure configuration and software deployment to map graphs. To do so, we map attacked services in corresponding vertices in BPDs, LEGs and ODGs so that the source graphs can be mapped to the destination graphs. A LEG is easily mapped to a BPD by matching the attacked services to the corresponding tasks supported by the services. We use depth first search (DFS) method to map an ODG to a LEG, which is shown in Algorithm 1.

In Algorithm 1, all object nodes in an ODG are initially marked as *having not been checked* by using color white as shown in the *for* loop between Line 2 and 4. Then, for each given unchecked object node V_O (Line 6 and 7), the algorithm repeatedly calls $Find(V_O, LEG)$ function (calls from Line 13, and the function itself is between Line 36 and Line 50), attempting to find the matching post-attack status node in the LEG by checking if the attacked service in the LEG is equal to the attacked service in the ODG. If such a post-attack status node, say N_{c1}, is found (Line 15), then, Algorithm 1 checks if the attack step between Node V_O and its parent node $parent(V_O)$ in the ODG has a mapping attack step between Node N_{c1} and its parent node(s) $parent(N_{c1})$ in the LEG (Line 19). If there is no such a matching attack step, one is added to the LEG (Line 22 to 25). If there

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

Algorithm 1: Mapping an ODG to a LEG

Input: An ODG=(V, V_E, E) and a LEG=(N_r, N_f, N_c, E, L, G).
Output: A LEG integrated with attack paths from the ODG.

```
//Color all nodes in ODG WHITE
for each node V_O in ODG do
    color[V_O] ← WHITE
end
//Go through each object in ODG
for each node V_O in ODG do
    if V_O == WHITE then
        //Initialize all nodes in LEG white
        for each node N_c in LEG do
            color[N_c] ← WHITE
        end
        //Search for the corresponding N_c1 in LEG
        N_c1 = Find(V_O, LEG)
        //If there is such a matching N_c1
        if N_c1 ≠ ∅ then
            color(V_O) ← BLACK
            //See if the object's parent matches
            //corresponding N_c1' s parent
            N_c2=Find(parent(V_O), LEG)
            //If not matching parents,
            //add the missing attack step from ODG to LEG
            if N_c2 ≠ parent(N_C1) then
                LEG ← Flow(N_c1, V_E)
                LEG ← Flow(V_E, N_c2)
            end
        end
        else
            //If there is no such a matching N_c1 in LEG
            //Add the new object to LEG
            LEG ← V_O
            color [V_O]=GRAY
        end
        V_O =child(V_O)
    end
end
Function Find(V_O, LEG)
    //Go through each N_c in LEG
    for each post attack status N_c from LEG do
        //Check if there is any matching N_c for V_O
        if (N_c.service == V_O.service AND
        color[N_c] == white then
            color[N_c] ← BLACK
            return N_c
        end
        else
            color[N_c] ← GRAY
            N_c ← the child post attack status node of N_c
        end
    end
    return ∅
```


Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

isn't a mapping post-attack status Node N_{c1} in the LEG for Node V_O (Line 27) in the ODG, one is added to the LEG (Line 27 to 32), and the search continues (Line 33) until all nodes in the ODG are checked (i.e. colored).

3.5 Computing Mission Impact

We propose to use the interval [0,1] to quantify a mission impact of an attack, computed by using the following steps.

3.5.1 Compute the impact scores of attacks in LEGs

In a LEG, we use $P(a)$ to represent the impact of attacks on services deployed on host computers. NIST's NVD-CVSS published reported vulnerabilities with assigned impact scores, which we propose to use for each P(a) if an attack a can be found in NIST's NVD. If the attack a cannot be found in NIST's NVD, we suggest using expert knowledge to assign an impact score to $P(a)$. We use our previous work [9] to compute a cumulative impact score of attacks on the same service as follows.

$$P(a) = P(a_1) \cup P(a_2) \tag{1}$$

In Equation 1, a_1 and a_2 are two attacks on the same service. $P(a_1) \cup P(a_2) = P(a_1) + P(a_2) - P(a_1) \times P(a_2)$.

3.5.2 Assign weight to tasks/missions

A value between [0,1] is proposed as the weight of mission impact of attacks on a task, indicating the importance of the corresponding task to the mission of a business process.

3.5.3 Compute mission attack impacts in BPDs

We map LEGs to BPDs so that the mission impact of attacks I(T) on a task T is computed using Equation 2.

$$I(T) = weight \times P(T) \tag{2}$$

$$P(T) = P(a) \tag{3}$$

$$P(T) = P(a_1) \cup P(a_2) \tag{4}$$

In Equation 2, P(T) is the impact of attacks on a task T in a BPD. Depending on the mapping relationship from the attacked service(s) (represented by a, a_1, a_2) in a LEG to a task (represented

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

by T) in a BPD, P(T) is computed by using Equation 3(one-to-one mapping relationship) and Equation 4(many-to-one mapping relationship) respectively.

3.5.4 Compute the cumulative mission impact

Mission impact of attacks on each task can be computed using Equation 2, Equation 3 and Equation 4. However, in some cases, the cumulative mission attack impact for the final mission is required to estimate the overall damage, which we compute as the maximum. We use M to represent the cumulative mission impact. Correspondingly, for the four kinds of relationships between tasks composing of a business process, M is computed using the following equations, explained below.

$$M(B) = Max\{I(T_1), I(T_2), \ldots, I(T_n)\} \quad (5)$$

$$M(B) = Max\{I(T_1), I(T_2), I(T_4) \ldots, I(T_n)\} \quad (6)$$

$$M(B) = Max\{I(T_1), I(T_3), I(T_4) \ldots, I(T_n)\} \quad (7)$$

$$M(B) = Max\{M(B_{before}), I(T_2\prime)\} \quad (8)$$

1. If the tasks $T_1, T_2, \ldots, T_n$ composing of the final mission B have a sequential relationship with each other, or among them, there are tasks, say T_2, T_3 that have a parallel fork relationships with the predecessor task T_1 and a parallel merge relationships with the successor tasks T_4, M(B) is computed as shown in Equation 5.

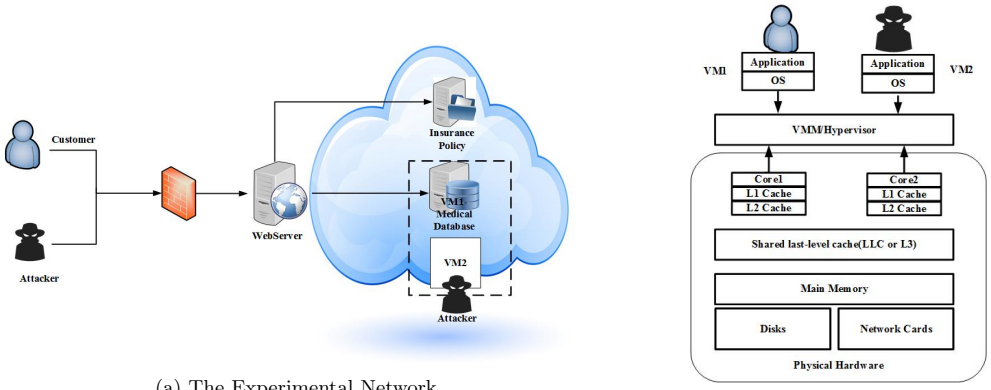
2. Among tasks $T_1, T_2, \ldots, T_n$ that compose of the final mission B, if there are tasks, say T_2, T_3, which have an exclusive decision relationship with the predecessor task T_1 and the successor tasks T_4, and all other tasks including $T_4, \ldots, T_n$ have a sequential relationship with each other, depending on which task (either T_2 or T_3) the business process chooses, either Equation 6 or Equation 7 is used to compute M(B).

3. Suppose tasks $T_1, T_2, \ldots, T_n$ compose of the final mission B in Pool 1. We use $M(B_{before})$ to represent the cumulative mission attack impact of B without any message passing from other pools. If there is message passing relationship between a task $T_2\prime$ from Pool 2 to T_2 in Pool 1, Equation 8 is used to compute M(B).

4 The Case Study

This section describes our case study used to show the utility of our model.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.



(a) The Experimental Network

(b) VM1 and VM2

Figure 2: The experimental network and attacks using cloud services

4.1 The experimental network and attacks

Figure 2a shows our experimental network configured to manage the customers' medical records and their health insurance policy files. These records and files are stored on two separate VMs in a private cloud set up by using OpenStack (we used the version *Juno 2014.2.3*). OpenStack is a collection of python-based software projects that manage access to pooled storage, computing and network resources that reside in one or many machines of a cloud system [17]. These projects include *Neutron (Networking)*, *Nova (Compute)*, *Glance (Image Management)*, *Swift (Object Storage)*, *Cinder (Block Storage)* and *Keystone (Authorization and Authentication)*. OpenStack can be used to deploy *SaaS*, *PaaS* and *IaaS* cloud models, but is mostly deployed as an *IaaS* cloud. Authenticated users can access the file server to retrieve policy files using *ssh* and query the medical database stored in the database server using *MySQL* queries through a web application.

We assume that the attacker's objective is to steal customers' medical records, prevent the medical records' availability and modify the health insurance policies. By probing the deployed web and cloud services, as the attacker, we launched the following attacks.

The SQL injection attack: Because our web application did not sanitize user input, the attacker could use it to create a SQL injection attack (CWE-89) to access customers' medical records. By using the query *select * from profile where name='Alice' and (password='alice' or '1' = '1')*, where *profile* was the database name, and *'1'='1'* was the payload that made the query bypass the password check, we retrieved all customer medical records.

The DoS attack: According to NIST's NVD, the vulnerability *CVE-2015–3241* that is in OpenStack Compute (Nova) versions *2015.1 through 2015.1.1, 2014.2.3 and earlier* allows authenticated users to cause Denial of Services (DoS) by re-sizing and then deleting an instance (VM). The process of resizing and deleting an instance is also called an instance migration. The migration process

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

does not terminate when an instance is deleted with *CVE-2015-3241*, so an authenticated user could bypass user quota enforcement to deplete all available disk space by repeatedly performing instance migration. By using this vulnerability, playing a privileged IaaS malicious user, we launched the DoS attack toward the database server by repeatedly re-sizing and deleting VM2 that co-resided in the same physical machine as the medical database server(VM1).

The cross-VM side-channel attack: Side-channel attacks can be used to extract fine grain information across VMs that reside on the same hypervisor [18].

In our experimental network, we simulated Yarom's cache side-channel attack [19] shown in Figure 2b on our two VMs that co-reside in the same multi-core processor (an Intel quad-core i7). In this attack, the cache shared between the victim and attacking VMs can be used to fill with data from the attacking machine. Each time when an encryption occurs, the processor evicts one or more lines of the attacker's memory from the shared cache, causing timing variation. By measuring the timing, we can obtain the information to hijack the encryption key being used with the GNU Privacy Guard (GnuPG) application in the victim VM. Because Yarom's attack uses the implementation weakness existing in GnuPG 1.x before 1.4.16 versions to obtain the information used to extract the private encryption key, we installed GnuPG 1.4.12 in our VM1 (the medical database server), and executed the attack from the VM2.

The social engineering attack: We simulated a social engineering attack toward the file server. Assuming that the attacker obtained a legitimate user's (username, password) credentials, the attacker could easily log into the file server, using the user's privilege to modify corresponding insurance policy files in the file server.

To capture attacks, in our experimental network, we deployed snort as the IDS, installed Wireshark to monitor network traffic, and configured all servers to log users' access. Also, in order to obtain evidence for those attacks that are missed by the IDS alerts and service logs, we intercepted system calls from the users' processes in the cloud.

4.2 The three levels of graphs

By using the network configuration, service deployment and captured evidence, we constructed the three levels of graphs, including a BPD, two LEGs and two ODGs, described as follows.

1. The Business Process Diagram

   Figure 3 shows our BPD with 3 pools. They are Pool 1 (Web Interface), Pool 2 (Public Cloud Service) and Pool 3 (IaaS User Service). The business process in Pool 1 is the web interface for the clients(medical customers), which is composed of start/end events, two consecutive tasks *enter username/password*, *send out request* and an exclusive gateway for tasks *review policy file*

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

and *review medical records* that depends on the client's request. The business process in Pool 2 is composed of start/end events, a task *check user request* followed by an exclusive gateway that directs to two tasks *request policy files* and *request customer databases*, which depends on the message passed from the task *send out request* in pool 1. In each of the decision task branch, there is an exclusion decision gateway(named *file available* and *data available* respectively) followed by tasks that either send the data(*policy file* or *customer medical records* through the message passing) back to the clients or reject the customer's requests otherwise. Pool 3 is a business process used to describe IaaS user services, which is mainly composed of three tasks *encrypt data in VM1*, *resize VM2*, and *install and run program in VM2* that are the exclusive decisions of the task *check IaaS user request*. For each of the three business processes in the three pools, we consider the last task(s) before the end event as the business process mission(s).

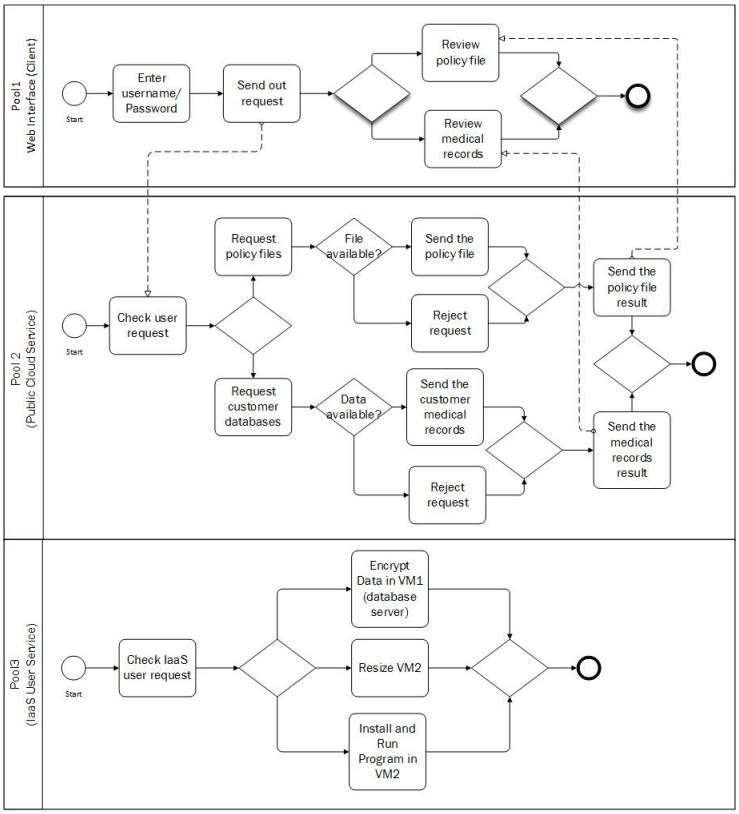


Figure 3: The BPD of the experimental network

2. The Logical Evidence Graph

Table 2 shows evidence of the SQL injection attack with Snort alerts and database access logs. Using timestamps, corresponding alert content and MySQL general query logs, we asserted that the attacker used a typical SQL injection with payload *'1'='1'* to attack the customers' database

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

Table 2: The snort alert and database server log of SQL injection attack

Time Stamp	Machine	IP Address/Port	Snort Alert and Database Server Access Log
	Attacker	129.174.124.122	
06/13-14:37:27	web server	129.174.124.184	SQL injection attack(CWE-89)
13/Jun/2017:14:37:34	Database server	129.174.124.35	Access from 129.174.124.184

in the database server. Our IDS failed in capturing the DoS attack launched by exploiting the vulnerability CVE-2015-3241 in OpenStack Nova services. Because OpenStack application programing interface (API) logs provide users' operations of running instances (we illustrate some of these API logs in Figure 4, where we use bold font to show the users' operations), we used them to conclude that the IaaS user in VM2 (the attacker in our experiment) kept re-sizing and deleting the instance VM2 that co-resided in the same physical machine as the database server (VM1), which caused the DoS attack toward the database server.

2017-07-18 07:52:00.237 DEBUG oslo_concurrency.processutils [req-f79c7911-04ed-4a0c-adbe-0ae0a487c0f7 admin admin] Running cmd (subprocess): **mv /opt/stack/data/nova/instances/bd1dac18-1c e2-44b5-93ee-967fec640ff3= /opt/stack/data/nova/instances/bd1dac18-1ce2-44b5-93ee-967fec640ff3_resize from (pid=41737)** execute /usr/local/lib/python2.7/dist-packages/oslo_concurrency/processutils.py:344

2017-07-18 07:52:00.253 DEBUG oslo_concurrency.processutils [req-f79c7911-04ed-4a0c-adbe-0ae0a487c0f7 admin admin] CMD **"mv /opt/stack/data/nova/instances/bd1dac18-1ce2-44b5-93ee-967fec640ff3 /opt/stack/data/nova/instances/bd1dac18-1ce2-44b5-93ee-967fec640ff3_resize"** returned: 0 in 0.016s from (pid=41737) execute /usr/local/lib/python2.7/dist-packages/oslo_concurrency/processutils.py:374

2017-07-18 07:52:00.254 DEBUG oslo_concurrency.processutils [req-f79c7911-04ed-4a0c-adbe-0ae0a487c0f7 admin admin] Running cmd (subprocess): **mkdir –p /opt/stack/data/nova/instances/bd1dac18-1ce2-44b5-93ee-967fec640ff3 from (pid=41737)** execute /usr/local/lib/python2.7/dist-packages/oslo_concurrency/processutils.py:344

Figure 4: OpenStack Nova API call logs

```
/* the initial attack location and final attack status*/
attackerLocated(internet).
attackGoal(execCode(database,user)).
/* the network access configuration*/
hacl(internet, webServer, tcp, 80).
hacl(webServer, database, tcp, 3306).
/* configuration information of webServer */
vulExists(webServer, 'directAccess', httpd).
vulProperty('directAccess', remoteExploit, privEscalation).
networkServiceInfo(webServer , httpd, tcp , 80 , apache).
/* the vulnerability of the web application */
vulExists(database, 'CWE-89', httpd).
vulProperty('CWE-89', remoteExploit, privEscalation).
networkServiceInfo(database , httpd, tcp , 3306, user).
```

Figure 5: Prolog predicates for SQL injection

In order to use the forensic analysis tool mentioned in Section 3.2, we converted system configu-

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

```
/* the initial attack status of being an iaas user and the final attack status*/
attackerLocated(iaas).
attackGoal(execCode(nova,admin)).
/*the cloud configuration, the "_" represents any protocol and port*/
hacl(iaas,nova,_, _).
/* the vulnerability in nova */
vulExists(nova, 'CVE-2015-3241', 'REST').
vulProperty('CVE-2015-3241',remoteExploit, privEscalation).
networkServiceInfo(nova , 'REST', http, _, admin).
```

Figure 6: Prolog predicates for DoS attack

rations and the evidence for the SQL injection attack and DoS attack to Prolog predicates shown in Figures 5, 6. The output LEGs produced by the tool are shown in Figure 7 and Figure 8 respectively with node names in Tables 3, 4. The two LEGs are not grouped together due to distinct attacker locations and privileges. Consider an example attack step (Nodes $3, 7, 8 \rightarrow 2 \rightarrow 1$ in Figure 8). Facts of LEGs are shown in Nodes 7, 8, modeling pre-attack configurations and vulnerabilities. The consequence fact node (Node 1) shows post-attack evidence derived by applying a rule (an ellipse Node 2 connecting Nodes 3, 7, 8 to the post attack status, Node 1) to the parent facts (Nodes 7, 8) and parent consequence fact (Node 3 that is obtained from a prior stepping stone step).

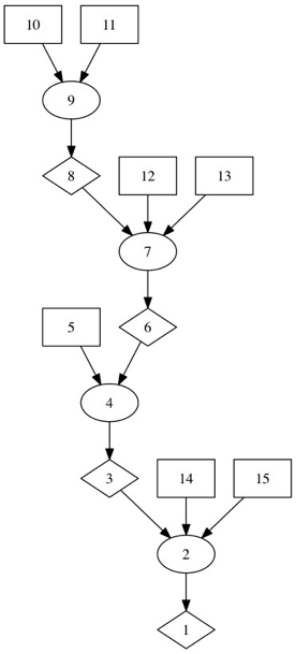


Figure 7: The LEG of SQL injection attack toward the database

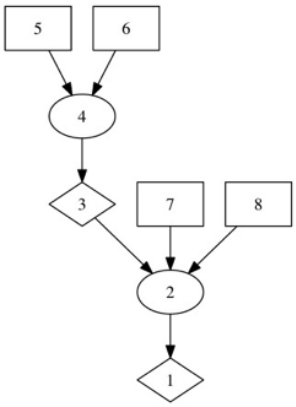


Figure 8: The LEG of DoS attack toward the database server

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

Table 3: Names of nodes in Figure 7

No.	Notation of all nodes
1	execCode(database,_)
2	RULE 2 (remote exploit of a server program)
3	netAccess(database,tcp,3306)
4	RULE 5 (multi-hop access)
5	hacl(webServer,database,tcp,3306)
6	execCode(webServer,apache)
7	RULE 2 (remote exploit of a server program)
8	netAccess(webServer,tcp,80)
9	RULE 6 (direct network access)
10	hacl(internet,webServer,tcp,80)
11	attackerLocated(internet)
12	networkServiceInfo(webServer,httpd,tcp,80,apache)
13	vulExists(webServer,'directAccess',httpd, remoteExploit,privEscalation)
14	networkServiceInfo(database,httpd,tcp,3306,_)
15	vulExists(database,'CWE-89',httpd,remoteExploit, privEscalation)

Table 4: Names of nodes in Figure 8

No.	Notation of all nodes
1	execCode(nova,admin)
2	RULE 2 (remote exploit of a server program)
3	netAccess(nova,http,_)
4	RULE 6 (direct network access)
5	hacl(cloud,nova,http,_)
6	attackerLocated(cloud)
7	networkServiceInfo(nova,'REST',http,_,admin)
8	vulExists(nova,'CVE-2015-3241', 'REST',remoteExploit,privEscalation)

3. The Object Dependency Graph

Due to the lack of IDS alerts and logs for the side-channel and social engineering attacks, we could not reconstruct the two attack scenarios in the form of LEGs as we did for the SQL injection and DoS attacks. Therefore, we turned to the corresponding system calls we captured during the attacks and used the method mentioned in Section 3.3 to construct ODGs for a forensic analysis.

Figure 9 and Figure 10 are a fraction of the system calls captured from VM1 (the database server) and VM2 (the attacker's VM) during the side-channel attack. The system call in Figure 9 shows that a file from VM1 (named *message.txt*) was encrypted by using GnuPG 1.4.12. System calls in Figure 10 show that a probe program *bin/probe* (Line 1) in VM2 used *mmap2* (Line 3)

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

to force the underlying system to share memory addresses (Lines 5, 6, 7...) with the probing process; hence the probing process could read data from the shared memory addresses between VM1 and VM2 for a later malicious analysis (Lines 10, 11, 12 and Line 1 show the data was written to a file named *out.txt*, and our continuous captured system calls show, later, a Python program was used to extract and analyze the information in *out.txt*). Figure 11 is a fraction of the system calls captured from the file server, where the *read/write* system call trace shows that the *test.txt* in *FileServer* has been modified. Because the corresponding sshd log in the *FileServer* recorded the users' access, it was easy to judge that the attacker stole a legitimate user's credentials to modify the policy file (the sshd log is omitted).

```
execve("/home/flush-reload-master/build_gpg/gnupg-1.4.12/bin/gpg", ["/home/flush-reload-"..., "--yes", "-
-sign", "message.txt"], [/* 61 vars */]) = 0
brk(0)                                  = 0x942b000
......
```

Figure 9: Filtered system calls of the side-channel attack from VM1

```
1.  execve("bin/probe", ["bin/probe", "/home/flush-reload-"..., "docs/addr/osx.txt", "out.txt", "5"], [/*
    61 vars */]) = 0
    …
2.  fstat64(4, {st_mode=S_IFREG|0664, st_size=78, ...}) = 0
3.  mmap2(NULL, 4096, PROT_READ|PROT_WRITE, MAP_PRIVATE|MAP_ANONYMOUS, -
    1, 0) = 0xb7702000
4.  read(4, "4\n0x00000000000967c7\n0x000000000"..., 4096) = 78
5.  write(1, "Probing 4 addresses:\n", 21)  = 21
6.  write(1, "0x967c7\n", 8)                = 8
7.  write(1, "0x95f5d\n", 8)                = 8
    …
8.  write(1, "Started spying\n", 15)        = 15
    …
9.  mmap2(NULL, 4096, PROT_READ|PROT_WRITE, MAP_PRIVATE|MAP_ANONYMOUS, -
    1, 0) = 0xb7701000
10. write(5, "0 0 22690\n0 1 460\n0 2 1159\n0 3 6"..., 4096) = 4096
11. write(5, " 1 217\n113 2 223\n113 3 214\n114 0"..., 4096) = 4096
12. write(5, "520\n215 3 232\n216 0 490\n216 1 26"..., 4096) = 4096
    …
```

Figure 10: Filtered system calls of the side-channel attack from VM2

Using the dependency rules listed in the left column of Table 1 and corresponding analysis on the system calls as shown in Figures 9, 10 and 11, we constructed two ODGs, showing the attacker from VM2 read the shared cache between VM1 and VM2, and the attacker from Internet used a legitimate user's credentials to access the file server and modified a policy file. We mapped both ODGs to the LEGs in Figure 7 and Figure 8. The integrated LEGs show that: (1) the attacker from the Internet launched two attacks including using stolen credentials to modify a policy file and stealing all customers' medical records by using a SQL injection attack; (2) the attacker who was an IaaS user launched two other attacks including a DoS attack and a side-channel attack to the database server.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

write(9, “v”, 1) = 1
read(11, “v”, 16384) = 1
write(3, “\0\0\0\20\331\255\275\264c\2173)z2j\32\255n\2007d\366m\21\316
\2648\240\207\31\211” . . . , 36) = 36
read(3,“\0\0\0\20\240\253\341\227\321xU\305\347\226\246\361\316\242S =
\30\341QT\231\n\343\314\343\307\f\361”. . . , 16384) = 36
write(9, “i”, 1) = 1
read(11, “i”, 16384) = 1
write(3,“\0\0\0\20\177\352\313\332\373yjM\3416l\230\215\10\220p\252g\375\365
\1\f\335\361\r\273\374\357”. . . , 36) = 36
read(3,“\0\0\0\20\27\334?\201x\300\16\356\346, \0379\32\220{\372)\366\4\v\1 =
\347\263\311\250k\353” . . . , 16384) = 36
write(9, “ ”, 1) = 1
read(11, “ ”, 16384) = 1
write(3,“\0\0\0\20ri\321\344\220\313\322\254S\252o\201\225; 6v\243\205\10gŝ
\253\237\325\375\332v” . . . , 36) = 36
read(3, “\0\0\0\20\5\27k; \254\301\24\n\\ZN\267\260\336\323ı\323\32\345\2b\
226 − \271|[B\21” . . . , 16384) = 36
write(9, “t”, 1) = 1
read(11, “t”, 16384) = 1
read(3,“\0\0\0\20\325\261\7\254\211(\201\331\272\344[\355\200\\u4\357G\347
\232\276 : \201\376\342\202\201.” . . . , 16384) = 36
write(3,“\0\0\0\20\320\254\#\312\211_\3022\n\227u\16I\372\202\347\37\252T
\257\220
\210E\343\222\342\24S” . . . , 36) = 36
write(9, “e”, 1) = 1
read(11, “e”, 16384) = 1
...
write(9, “\t”, 1) = 1
read(11, “st.txt ”, 16384) = 7
. . .

Figure 11: Filtered system calls of modifying a file from the file server

Table 5: The CVSS impact scores

Attack Name	**CVE Entry**	**Symbol Representation**	**Attack Impact**
SQL injection	CWE-89	N_1	0.9
DoS attack	CVE-2015-3241	$N_1\prime$	0.69
Social Engineering		N_s	0.5
side-channel attack	CVE-2013-4576	N_{sc}	0.29

4.3 Mission impact computation in our case study

The impact score of each attack step in all LEGs and ODGs is shown in Table 5, where the three impact scores of *CWE-89*, *CVE-2015-3241*, *CVE-2013-4576* were obtained from NIST's NVD-CVSS and the impact score of the social engineering attack was from our expert knowledge. The impact scores in NIST's NVD-CVSS are based on a [0, 10] scale, which we converted to a [0,1] interval scale.

We mapped all attacks from LEGs and ODGs to the BPD in Figure 3 and computed the mission impacts of the four attacks as shown in Table 6. Based on Table 6 and the BPD in Figure 3, the cumulative mission impacts are computed and listed in Table 7.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

Table 6: The mission impact scores

Pool	Task	Mapping Attack	Weight	Mission Impact
Pool 1	Check username and password	CWE-89	1	$I_1 = 1 \times P(N_1) = 1 \times 0.9 = 0.9$
Pool 2	Data available	CVE-2015-3241	0.9	$I_2 = 0.9 \times P(N_1\prime) = 0.9 \times 0.69 = 0.621$
Pool 2	Request policy files	Social Engineering	1	$I_3 = 1 \times P(N_s) = 1 \times 0.5 = 0.5$
Pool 3	Encrypt data in VM1	CVE-2013-4576	1	$I_4 = 1 \times P(N_{sc}) = 1 \times 0.29 = 0.29$

Table 7: The cumulative mission impact

Pool	Mission	Cumulative Mission Impact
Pool 1	Review policy file	M =Max(I_3) = $Max(0.5) = 0.5$
Pool 1	Review medical records	M = Max(I_1, I_2) = $Max(0.9, 0.621) = 0.9$
Pool 3	Encrypt data in VM1	M = Max(I_4) = $Max(0.29) = 0.29$

4.4 Using mission impacts to reduce attack risks

In a complex infrastructure, different missions use connections and combinations of multiple services. Each service is supported by software and hardware assets that are usually the target of attackers. In such cases, a tool can determine the impacts of cyber-attacks on the missions. By correlating the attacks on lower level assets to the higher level business process diagram and using mission impact scores of those attacks provided by NIST's NVD-CVSS, our model shows attacks and computes their impacts on complex missions. The information provided by such an analysis can be used by forensic investigators and infrastructure system planners.

As an example, we show how the mission impacts can be used by the system planners to enhance our experimental network. First, the cumulative impact scores in Table 7 show that the attacks on the mission of *review medical records* have a higher impact because the customers' medical records could be stolen by using a SQL injection attack. This suggests the user input sanitization should be implemented to defeat SQL injections. Second, the attacks and their impact scores shown in Table 6 show the two attacks (a DoS attack and a side-channel attack) were caused by cloud services that have vulnerabilities; hence the corresponding services should be moved to a more stable cloud that has countermeasures to attacks that can be launched through the shared hypervisor or host. Third, Table 6 shows, though the mission impact on the insurance policy stored in the file sever might not be as bad as the SQL injection attack on the database server, it has a score "0.5" that can not be neglected. An easy solution would be limiting the file *write/modify* right to only administrators with local access.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

5 Conclusion

We proposed a three-layered graphical model to quantify mission impacts of cyber-attacks in this work. We did so by reconstructing attacks based on available evidence from attack logs and system call sequences when logs missed requisite evidence to reconstruct attack steps. We used impact scores published in the NIST's NVD-CVSS and expert opinions when such numbers are unavailable as a base line to estimate attack impacts. We then mapped the attacks to higher-level business processes and considered their importance weight for business processes to compute the impacts of cyber-attacks on missions. Our case study showed that this model can be used to mitigate the impacts of cyber-attacks in a network. In the future, we will conduct more experiments using attacks on the cloud infrastructure to determine how our framework should be used in order to help enterprises to reduce the security risks of their infrastructures.

DISCLAIMER

This paper is not subject to copyright in the United States. Commercial products are identified in order to adequately specify certain procedures. In no case does such an identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the identified products are necessarily the best available for the purpose.

References

[1] S. Musman and A. Temin, A cyber mission impact assessment tool, In Technologies for Homeland Security (HST), 2015 IEEE International Symposium on 2015 Apr 14 (pp. 1-7).

[2] NIST National Vulnerability Database Common Vulnerability Scoring System, available at https://nvd.nist.gov/vuln-metrics/cvss.

[3] P. Mell and T. Grance. "NIST definition of cloud computing". National Institute of Standards and Technology. October 7, 2009.

[4] K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie. "Cloud forensics". In IFIP International Conference on Digital Forensics, pp. 35-46. Springer Berlin Heidelberg, 2011.

[5] S. Noel, J. Ludwig, P. Jain, D. Johnson, R. K. Thomas, J. McFarland, B. King, S. Webster and B. Tello, Analyzing mission impacts of cyber actions (AMICA), In NATO IST-128 Workshop on Cyber Attack Detection, Forensics and Attribution for Assessment of Mission Impact, Istanbul, Turkey, 2015.

[6] X. Sun, A. Singhal, P. Liu, Towards Actionable Mission Impact Assessment in the Context of Cloud Computing, In Livraga G., Zhu S. (eds) Data and Applications Security and Privacy XXXI. DBSec 2017. Lecture Notes in Computer Science, vol 10359.

[7] C. Liu, A. Singhal and D. Wijesekara, A logic-based network forensic model for evidence analysis, in Advances in Digital Forensics XI, G. Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 129-145, 2015.

[8] X. Sun, J. Dai, A. Singhal, P. Liu and J. Yen, Towards Probabilistic Identification of Zero-day Attack Paths, Accepted for IEEE Conference on Communication and Network Security, Philadelphia, October 17th 19th, 2016.

[9] C. Liu, A. Singhal and D. Wijesekera, Mapping evidence graphs to attack graphs, In Information Forensics and Security (WIFS), 2012 IEEE International Workshop on (pp. 121-126). IEEE.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

[10] Y. Sun, T. Y. Wu, X. Liu, X. and M.S. Obaidat, Multilayered Impact Evaluation Model for Attacking Missions, IEEE Systems Journal, 10(4), pp.1304-1315, 2016.

[11] X. Ou, S. Govindavajhala, S. and A. W. Appel, MulVAL: A Logic-based Network Security Analyzer, In USENIX Security Symposium (pp. 8-8), July 2005.

[12] S. Jajodia and S. Noel, Topological vulnerability analysis, In Cyber situational awareness, pp. 139-154. Springer US, 2010.

[13] W. Wang, E.D. Thomas, A graph based approach toward network forensics analysis, ACM Transactions on Information and Systems Security 12 (1) 2008.

[14] C. Liu, A. Singhal and D. Wijesekera, A Probabilistic Network Forensic Model for Evidence Analysis, IFIP International Conference on Digital Forensics. Springer International Publishing, 2016.

[15] Online Resource for Markup Language Technologies, retrieved from http://xml.coverpages.org/bpm.html#bpmi.

[16] L. Herbert, Specification, Verification and Optimization of Business Processes, A Unified Framework, Technical University of Denmark (2014).

[17] OpenStack Open Source Cloud Computing Software. Retrieved from https://www.openstack.org.

[18] Y. Zhang, A. Juels, M. K. Reiter and T. Ristenpart, Cross-vm side channels and their use to extract private keys, In Proceedings of the 2012 ACM Conference on Computer and Communications Security (New York, NY, USA, 2012), CCS '12, ACM, pp. 305–316.

[19] Yarom, Yuval, and Katrina Falkner. "FLUSH+ RELOAD: A High Resolution, Low Noise, L3 Cache Side-Channel Attack." In USENIX Security Symposium, pp. 719-732. 2014.

Liu, Changwei; Singhal, Anoop; Wijesekera, Duminda.
"A Layered Graphical Model for Cloud Forensic and Mission Impact Analysis."
Paper presented at 14th IFIP International Conference on Digital Forensics, New Delhi, India. January 3, 2018 - January 5, 2018.

Quantifying Variance Components for Repeated Scattering-Parameter Measurements*

Amanda Koepke and Jeffrey A. Jargon

National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305 USA
Email: amanda.koepke@nist.gov, jeffrey.jargon@nist.gov

***Abstract* — We quantify random uncertainties for scattering-parameters repeatedly measured with a vector network analyzer, focusing on variations due to multiple calibrations, disconnects, and repeat measurements. We describe a two-stage nested design, which allows us to model the random effects, and present results for a series of coaxial measurements performed by making use of an open-short-load-thru calibration kit with Type-N connectors.**

***Index Terms* — calibration, disconnect, measurement, repeat, scattering-parameters, two-stage nested design, variance.**

I. Introduction

Uncertainties of calibrated scattering parameters (*S*-parameters) measured with a vector network analyzer (VNA) can be classified as either random or systematic. In previous publications, we have quantified systematic uncertainties resulting from uncertainties in the physical models of our calibration standards [1-2]. In this study, we focus on random uncertainties, specifically on variations due to calibration, disconnect, and repeat measurements.

To quantify these variance components, we set up a two-stage nested, or hierarchical, design [3], depicted in Figure 1. Nesting refers to the imposed structure of the design: within each calibration we have disconnects specific to that calibration, and similarly within each disconnect we have repeat measurements specific to that disconnect. Repeat measurements are taken without breaking any connections, while a disconnect refers to disconnecting and reconnecting the cables before taking a new set of measurements.

In the following sections, we describe the two-stage nested design in detail, and present results for a series of coaxial measurements performed by making use of an open-short-load-thru (OSLT) calibration kit with Type-N connectors.

II. Two-Stage Nested Design

We assume the following random effects model:

$$\boldsymbol{Y}_{ijk} = \boldsymbol{\mu} + \boldsymbol{C}_i + \boldsymbol{D}_{(i)j} + \boldsymbol{\varepsilon}_{(ij)k}\,. \qquad (1)$$

Here $\boldsymbol{Y}_{ijk}$ is the multivariate response vector (i.e. calibrated magnitudes and phases of S_{21} measurements) under calibration i (i= 1, …, I), disconnect j (j= 1, …, J), and repeat k (k= 1, …, K). These vectors have dimension F×1, where F is the number of measured frequency points. The mean response $\boldsymbol{\mu}$ is a constant, and is the expected value of $\boldsymbol{Y}_{ijk}$ over all the calibrations, disconnects, and repeats. We assume the calibration effects $\boldsymbol{C}_i$ are random variables with mean $\vec{\mathbf{0}}$ and covariance matrix $\boldsymbol{\Sigma}_C$, the disconnect effects $\boldsymbol{D}_{(i)j}$ are random variables with mean $\vec{\mathbf{0}}$ and covariance $\boldsymbol{\Sigma}_D$, and the measurement errors $\boldsymbol{\varepsilon}_{(ij)k}$ are random variables with mean $\vec{\mathbf{0}}$ and covariance $\boldsymbol{\Sigma}$. All these effects are assumed to be independent. The notations *(i)j* and *(ij)k* denote nesting. The diagonal elements of the covariance matrices $\boldsymbol{\Sigma}_C$, $\boldsymbol{\Sigma}_D$, and $\boldsymbol{\Sigma}$ are the variance component vectors $\boldsymbol{\sigma}_C^2$, $\boldsymbol{\sigma}_D^2$, and $\boldsymbol{\sigma}^2$.

We are interested in the variance of the overall mean, $\bar{\boldsymbol{Y}}$…, and the variance components vectors $\boldsymbol{\sigma}_C^2$, $\boldsymbol{\sigma}_D^2$, and $\boldsymbol{\sigma}^2$. These can be estimated using a Multivariate Analysis of Variance (MANOVA) method [4]. Using this approach, we partition the sums of squares and cross products (SSCP) into three parts, $\mathbf{SSCP}_{Total}$=$\mathbf{SSCP}_C$+ $\mathbf{SSCP}_D$+$\mathbf{SSCP}_e$, where $\mathbf{SSCP}_C$, $\mathbf{SSCP}_D$, and $\mathbf{SSCP}_e$ are the sums of squares and cross products due to calibration, disconnect, and error. We use these to calculate the estimated mean square errors (MSE) due to the different factors, and from the MSE we estimate the variance of the overall mean and the variance components $\hat{\boldsymbol{\sigma}}_C^2$, $\hat{\boldsymbol{\sigma}}_D^2$, and $\hat{\boldsymbol{\sigma}}^2$.

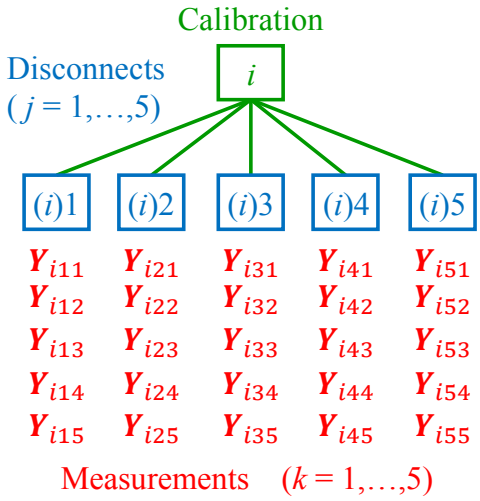


Figure 1. The two-stage nested design consisting of I calibrations, each including J disconnects and K repeated measurements taken at each disconnect. In this figure, J=K=5 for the i^{th} calibration.

Koepke, Amanda; Jargon, Jeffrey.
"Quantifying Variance Components for Repeated Scattering-Parameter Measurements."
Paper presented at 90th ARFTG Microwave Measurement Symposium, Boulder, CO, United States. November 30, 2017 - December 1, 2017.

To calculate the SSCP matrices, we first estimate the group means. For each disconnect $j = 1,2,\dots,J$ and calibration $i = 1,2,\dots,I$, we take the means of all repeat measurements

$$\overline{\boldsymbol{Y}}_{ij\cdot} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{Y}_{ijk}\,. \tag{2}$$

Next, for each calibration $i = 1,2,\dots,I$, we calculate the means of all J disconnects

$$\overline{\boldsymbol{Y}}_{i\cdot\cdot} = \frac{1}{J}\sum_{j=1}^{J} \overline{\boldsymbol{Y}}_{ij\cdot}\,. \tag{3}$$

Finally, the overall mean is calculated as

$$\overline{\boldsymbol{Y}}_{\dots} = \frac{1}{I}\sum_{i=1}^{I} \overline{\boldsymbol{Y}}_{i\cdot\cdot}\,. \tag{4}$$

Using these means, we calculate the sums of squares and cross products due to calibration ($\mathbf{SSCP}_C$), disconnect ($\mathbf{SSCP}_D$), and error ($\mathbf{SSCP}_e$).

For the error, we have:

$$\mathbf{SSCP}_e = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\big(\boldsymbol{Y}_{ijk} - \overline{\boldsymbol{Y}}_{ij\cdot}\big)\big(\boldsymbol{Y}_{ijk} - \overline{\boldsymbol{Y}}_{ij\cdot}\big)^T \tag{5}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\boldsymbol{Y}_{ijk}\boldsymbol{Y}_{ijk}^T - K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{\boldsymbol{Y}}_{ij\cdot}\overline{\boldsymbol{Y}}_{ij\cdot}^T, \tag{6}$$

which looks at deviations of individual measurements from their calibration/disconnect level group mean. This matrix has dimension $F\times F$. Similarly, for disconnect and calibration we have:

$$\mathbf{SSCP}_D = K\sum_{i=1}^{I}\sum_{j=1}^{J}\big(\overline{\boldsymbol{Y}}_{ij\cdot} - \overline{\boldsymbol{Y}}_{i\cdot\cdot}\big)\big(\overline{\boldsymbol{Y}}_{ij\cdot} - \overline{\boldsymbol{Y}}_{i\cdot\cdot}\big)^T \tag{7}$$

$$= K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{\boldsymbol{Y}}_{ij\cdot}\overline{\boldsymbol{Y}}_{ij\cdot}^T - JK\sum_{i=1}^{I}\overline{\boldsymbol{Y}}_{i\cdot\cdot}\overline{\boldsymbol{Y}}_{i\cdot\cdot}^T \tag{8}$$

and

$$\mathbf{SSCP}_C = JK\sum_{i=1}^{I}(\overline{\boldsymbol{Y}}_{i\cdot\cdot} - \overline{\boldsymbol{Y}}_{\dots})(\overline{\boldsymbol{Y}}_{i\cdot\cdot} - \overline{\boldsymbol{Y}}_{\dots})^T \tag{9}$$

$$= JK\sum_{i=1}^{I}\overline{\boldsymbol{Y}}_{i\cdot\cdot}\overline{\boldsymbol{Y}}_{i\cdot\cdot}^T - IJK\cdot\overline{\boldsymbol{Y}}_{\dots}\overline{\boldsymbol{Y}}_{\dots}^T\,. \tag{10}$$

For computational ease and notational clarity, we just consider the f^{th} diagonal elements of these matrices, as follows:

$$\mathrm{SS}_e(f) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{ijk}^2(f) - K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{Y}_{ij\cdot}^2(f), \tag{11}$$

$$\mathrm{SS}_D(f) = K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{Y}_{ij\cdot}^2(f) - JK\sum_{i=1}^{I}\overline{Y}_{i\cdot\cdot}^2(f), \tag{12}$$

and

$$\mathrm{SS}_C(f) = JK\sum_{i=1}^{I}\overline{Y}_{i\cdot\cdot}^2(f) - IJK\overline{Y}_{\dots}^2(f). \tag{13}$$

The associated f^{th} diagonal elements of the mean square error matrices are:

$$\mathrm{MS}_e(f) = \frac{\mathrm{SS}_e(f)}{IJ(K-1)}, \tag{14}$$

$$\mathrm{MS}_D(f) = \frac{\mathrm{SS}_D(f)}{I(J-1)}, \tag{15}$$

and

$$\mathrm{MS}_C(f) = \frac{\mathrm{SS}_C(f)}{I-1}. \tag{16}$$

The expected values of these mean squares are given in Table 1. By equating these expected mean squares with their counterparts estimated from the data and solving for the variance components $\boldsymbol{\sigma}^2$, $\boldsymbol{\sigma}_D^2$ and $\boldsymbol{\sigma}_C^2$, we obtain their estimates $\widehat{\boldsymbol{\sigma}}^2$, $\widehat{\boldsymbol{\sigma}}_D^2$, and $\widehat{\boldsymbol{\sigma}}_C^2$.

Notice the expected mean square error for calibration includes variability due to error, disconnect, and calibration, and since the variance components are nonnegative, $\mathrm{E}(\mathrm{MS}_C(f)) \geq \mathrm{E}(\mathrm{MS}_D(f)) \geq \mathrm{E}(\mathrm{MS}_e(f))$. However, since we are using mean squares estimated from the data to obtain our variance components estimates, there is a chance that $\mathrm{MS}_C(f) < \mathrm{MS}_D(f)$ or $\mathrm{MS}_D(f) < \mathrm{MS}_e(f)$, resulting in a negative variance component estimate. When this happens, we set the estimate equal to zero [5].

We can also use the mean squares to calculate the variance of the overall mean. Typically, the diagonal entries of the covariance matrix for $\overline{\boldsymbol{Y}}_{\dots}$ are calculated as:

$$\widehat{\mathrm{Var}\big(\overline{Y}_{\dots}(f)\big)} = \frac{\mathrm{MS}_C(f)}{IJK}. \tag{17}$$

However, due to the possibility of obtaining negative estimates for the variance components, this may underestimate the variance. This would occur when one or more of the variance components is close to zero. If there is no variability due to disconnect or calibration, $\mathrm{E}(\mathrm{MS}_C(f)) = \mathrm{E}(\mathrm{MS}_D(f)) = \mathrm{E}(\mathrm{MS}_e(f)) = \sigma^2$. In this case, all three estimated mean squares are estimates of the same quantity of interest, and it makes sense to use a weighted average of these terms, with the degrees of freedom as the weights to obtain the variance of the overall average:

TABLE I

Analysis of Variance Table. All formulas are written in terms of a single frequency, f, so SS_C, SS_D, and SS_e are the f^{th} diagonal elements of the sums of squares and cross-products (SSCP) matrices $\mathbf{SSCP}_C$, $\mathbf{SSCP}_D$, and $\mathbf{SSCP}_e$.

Source of Variation	Mean Squares (MS) – Estimated from the Data	Expected Mean Squares	Variance Component Estimates
Calibration	$\mathrm{MS}_C = \frac{\mathrm{SS}_C}{I-1}$	$\mathrm{E}(\mathrm{MS}_C) = \sigma^2 + K\sigma_D^2 + JK\sigma_C^2$	$\hat{\sigma}_C^2 = \frac{\mathrm{MS}_C - \mathrm{MS}_D}{JK}$
Disconnect	$\mathrm{MS}_D = \frac{\mathrm{SS}_D}{I(J-1)}$	$\mathrm{E}(\mathrm{MS}_D) = \sigma^2 + K\sigma_D^2$	$\hat{\sigma}_D^2 = \frac{\mathrm{MS}_D - \mathrm{MS}_e}{K}$
Error	$\mathrm{MS}_e = \frac{\mathrm{SS}_e}{IJ(K-1)}$	$\mathrm{E}(\mathrm{MS}_e) = \sigma^2$	$\hat{\sigma}^2 = \mathrm{MS}_e$

$$
\begin{aligned}
&\mathrm{Var}\big(\widehat{\bar{Y}_{\cdot\cdot}}(f)\big) \\
&= \frac{1}{IJK}\frac{(I-1)\mathrm{MS}_C(f) + I(J-1)\mathrm{MS}_D(f) + IJ(K-1)\mathrm{MS}_e(f)}{(I-1)+I(J-1)+IJ(K-1)} \\
&= \frac{\mathrm{SS}_C(f) + \mathrm{SS}_D(f) + \mathrm{SS}_e(f)}{IJK(IJK-1)} = \frac{\mathrm{SS}_{Total}(f)}{IJK(IJK-1)} \\
&= \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\big(Y_{ijk}(f) - \bar{Y}_{\cdot\cdot\cdot}(f)\big)^2}{IJK(IJK-1)}.
\end{aligned}
\tag{18}
$$

The last term is recognizable as the formula for the standard error of the mean. In theory, it is best to test if the individual variance components are zero, but in practice it is easiest to simply calculate the variance estimates from both Equation (17) and (18) and use whichever is largest, which is what we report. The possibility of obtaining negative estimates of variance components is a major drawback of this approach. This issue could be avoiding by using a different estimation procedure that would not allow for negative estimates, such as restricted maximum likelihood [3, 5].

The formulas presented here apply for a balanced experimental design, which means the number of observations within each treatment are the same. Here, the term 'treatment' refers to calibration and disconnect. For an unbalanced design, these formulas would have to be modified [3]. There are advantages to using a balanced design [4], especially when testing for differences between treatments, and for ease of computation, hence our use of a balanced experimental design.

III. Measurements

We examined variance components by performing four calibrations (I=4) using a single Open-Short-Load-Thru (OSLT) calibration kit with Type-N coaxial connectors. Physical models of the calibration standards were developed and validated with a multiline Thru-Reflect-Line (TRL) calibration within the NIST Microwave Uncertainty Framework [2]. Within each calibration, we connected and disconnected a 20-dB attenuator five times (J=5), and made repeated measurements five times (K=5) during each connection, as illustrated in Figure 1. Thus, the attenuator was measured 100 (4×5×5) times. All measurements were performed on a frequency grid from 0.2-18 GHz in steps of 0.2 GHz (360 points).

Figures 2 and 3 plot the overall means ($\bar{\boldsymbol{Y}}_{\cdot\cdot\cdot}$) of the attenuator's measured magnitude and detrended phases of the transmission coefficient ($|S_{21}|$ and $\mathrm{Arg}\{S_{21}\}$). These two figures also display the means of the five disconnects and five repeat measurements for each of the four calibrations ($\bar{\boldsymbol{Y}}_{1\cdot\cdot}, \bar{\boldsymbol{Y}}_{2\cdot\cdot}, \bar{\boldsymbol{Y}}_{3\cdot\cdot}$, and $\bar{\boldsymbol{Y}}_{4\cdot\cdot}$). Variations among the calibrations are clearly visible.

Figures 4 and 5 plot the square root of variance component estimates, which correspond to the variation due to error ($\hat{\boldsymbol{\sigma}}$), the variation due to disconnect ($\hat{\boldsymbol{\sigma}}_D$), and the variation due to calibration ($\hat{\boldsymbol{\sigma}}_C$). For both $|S_{21}|$ and $\mathrm{Arg}\{S_{21}\}$, the variation due to error, $\hat{\boldsymbol{\sigma}}$, is negligible compared to the variation due to disconnect and calibration, $\hat{\boldsymbol{\sigma}}_D$ and $\hat{\boldsymbol{\sigma}}_C$. For $|S_{21}|$, the maximum value of $\hat{\boldsymbol{\sigma}}_D$ is 0.11 dB and the maximum value of $\hat{\boldsymbol{\sigma}}_C$ is 0.13 dB. For $\mathrm{Arg}\{S_{21}\}$ the maximum value of $\hat{\boldsymbol{\sigma}}_D$ is 0.48 degrees and the maximum value of $\hat{\boldsymbol{\sigma}}_C$ is 0.52 degrees. Surprisingly, these estimates do not gradually increase with frequency, but rather rise and fall unexpectedly within the measured frequency range. Furthermore, neither the values of $\hat{\sigma}_D(f)$ nor $\hat{\sigma}_C(f)$ consistently dominate the total variability throughout the measured frequency range.

Figures 6 and 7 plot the uncertainties due to random and systematic effects for the attenuator's $|S_{21}|$ and $\mathrm{Arg}\{S_{21}\}$ values. The random portion was calculated as the square root of the maximum of eq. (17) or (18), so $\sqrt{\mathrm{Var}\big(\widehat{\bar{Y}_{\cdot\cdot\cdot}}(f)\big)}$. The systematic portion was determined by the Microwave Uncertainty Framework. The figures illustrate that the magnitude of the uncertainty due to random effects is on the same order as the uncertainty due to systematic effects at frequencies below about 8 GHz, while at higher frequencies the uncertainty due to systematic effects is larger. For $|S_{21}|$, the maximum value of the uncertainty due to systematic effects is 0.21 dB, and for $\mathrm{Arg}\{S_{21}\}$, the maximum value is 1.79 degrees.

Koepke, Amanda; Jargon, Jeffrey.
"Quantifying Variance Components for Repeated Scattering-Parameter Measurements."
Paper presented at 90th ARFTG Microwave Measurement Symposium, Boulder, CO, United States. November 30, 2017 - December 1, 2017.

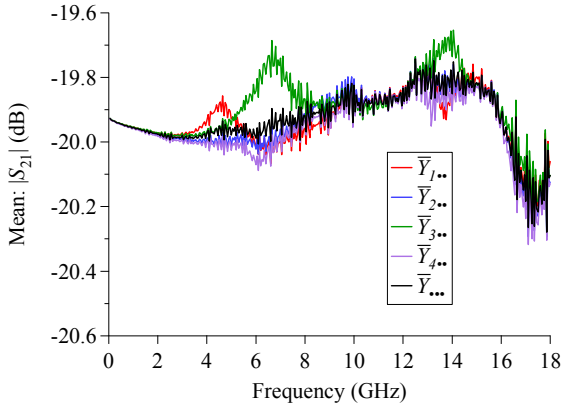


Fig. 2. Overall mean of the attenuator's $|S_{21}|$ values, and means of the J disconnects and K repeats for each i^{th} calibration.

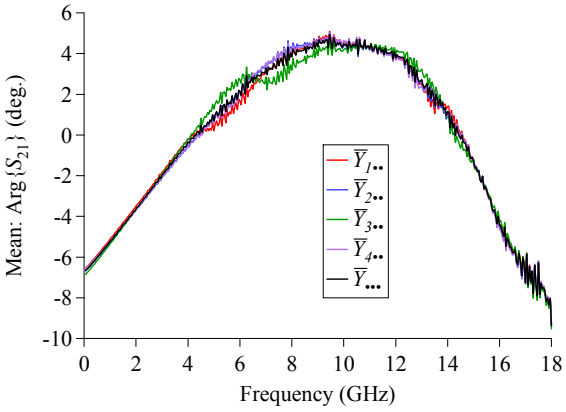


Fig. 3. Overall mean of the attenuator's Arg$\{S_{21}\}$ values, and means of the J disconnects and K repeats for each i^{th} calibration.

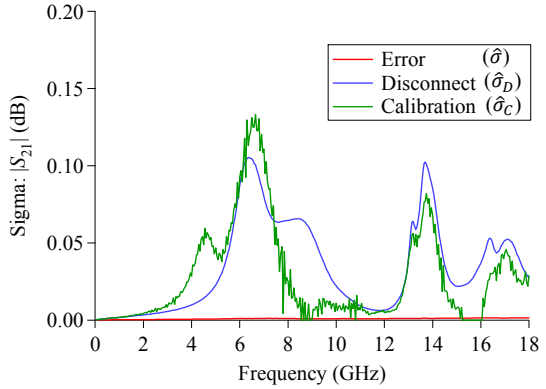


Fig. 4. Square root of variance component estimates for the attenuator's $|S_{21}|$ values including error $(\hat{\sigma})$, disconnect $(\hat{\sigma}_D)$, and calibration $(\hat{\sigma}_C)$.

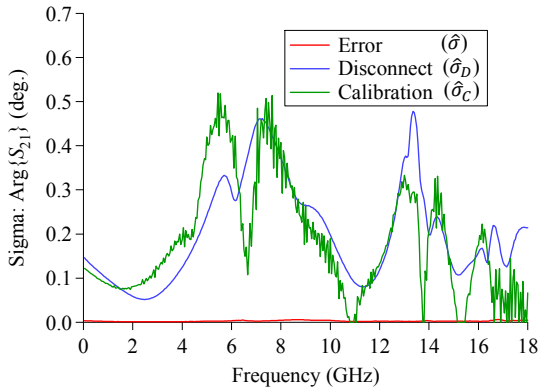


Fig. 5. Square root of variance component estimates for the attenuator's Arg$\{S_{21}\}$ values, including error $(\hat{\sigma})$, disconnect $(\hat{\sigma}_D)$, and calibration $(\hat{\sigma}_C)$.

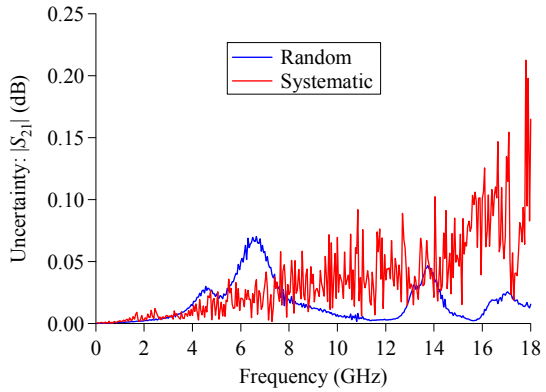


Fig. 6. Uncertainties due to random and systematic effects for the attenuator's $|S_{21}|$ values, calculated from repeat measurements (random) and the Microwave Uncertainty Framework (systematic).

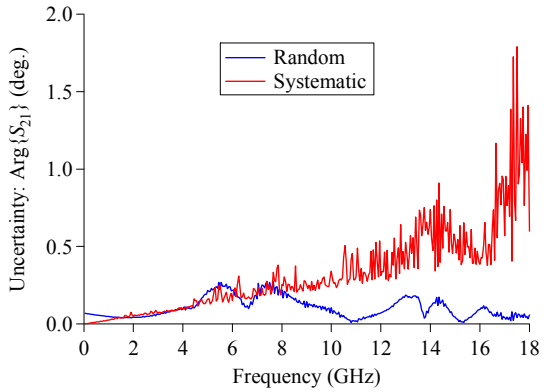


Fig. 7. Uncertainties due to random and systematic effects for the attenuator's Arg$\{S_{21}\}$ values, calculated from repeat measurements (random) and the Microwave Uncertainty Framework (systematic).

Koepke, Amanda; Jargon, Jeffrey.
"Quantifying Variance Components for Repeated Scattering-Parameter Measurements."
Paper presented at 90th ARFTG Microwave Measurement Symposium, Boulder, CO, United States. November 30, 2017 - December 1, 2017.

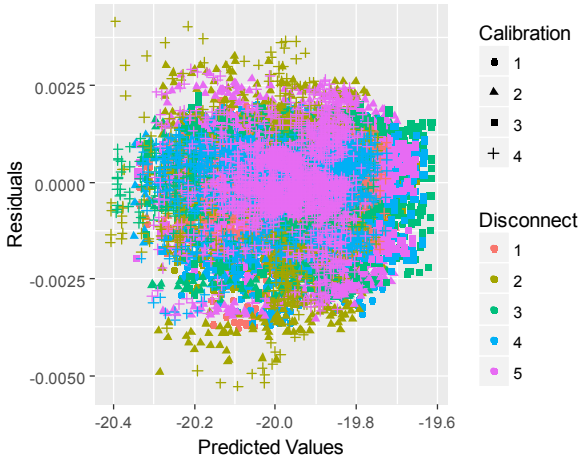


Fig. 8. Residuals versus predicted values for the attenuator's $|S_{21}|$ values. There is no obvious pattern in the residuals and there do not appear to be any outliers.

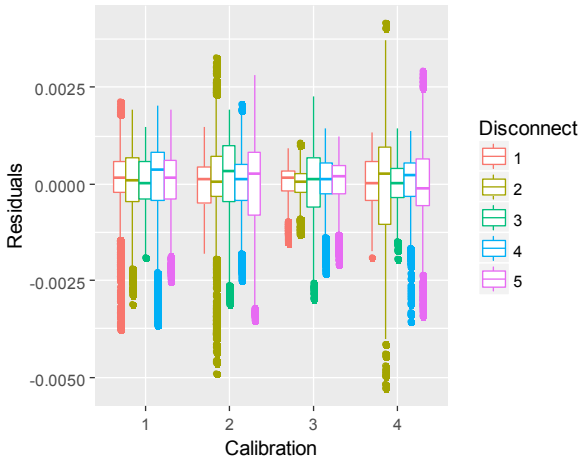


Fig. 9. Boxplots of residuals for the attenuator's $|S_{21}|$ values, separated by calibration and colored by disconnect. The spread of the points, summarized by the boxplots, seems comparable among the different disconnects and calibrations.

We should note that it is important to check for any possible violations of the assumption that our data follows the random-effects model described in eq. 1. To do this, we calculate the residuals, $\boldsymbol{e}_{ijk} = \boldsymbol{Y}_{ijk} - \widehat{\boldsymbol{Y}}_{ijk}$, where $\widehat{\boldsymbol{Y}}_{ijk}$ is the predicted value of $\boldsymbol{Y}_{ijk}$. Formally $\widehat{\boldsymbol{Y}}_{ijk} = \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{C}}_i + \widehat{\boldsymbol{D}}_{(i)j}$, where $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{Y}}_{\cdots}$, $\widehat{\boldsymbol{C}}_i = \overline{\boldsymbol{Y}}_{i\cdot\cdot} - \overline{\boldsymbol{Y}}_{\cdots}$, and $\widehat{\boldsymbol{D}}_{(i)j} = \overline{\boldsymbol{Y}}_{ij\cdot} - \overline{\boldsymbol{Y}}_{i\cdot\cdot}$ for the two-stage nested design. Thus $\widehat{\boldsymbol{Y}}_{ijk} = \overline{\boldsymbol{Y}}_{ij\cdot}$, giving $\boldsymbol{e}_{ijk} = \boldsymbol{Y}_{ijk} - \overline{\boldsymbol{Y}}_{ij\cdot}$. Plotting these residuals versus the predicted values, as shown in Figure 8 for the attenuator's $|S_{21}|$ measurements, allows us to look for any patterns that might suggest nonlinearity or points that might be outliers. These assumptions do not appear to be violated, as this plot resembles a random cloud of points with no obvious pattern in the residuals. Additionally, for the random effects model in eq. 1, we assume that the within-calibration variability is the same across the four calibrations. This can be checked with a plot of the residuals versus calibration, shown in Figure 9. We use boxplots of the residuals to summarize the many residuals, and separate by disconnect. A boxplot displays the distribution of a set of points. The bottom and top of the box denote the 25th and 75th percentiles of the data, so 50% of the data falls between these lines. The line in the box denotes the median (the 50th percentile). The thin vertical line and the points show the spread of the rest of the data. The points denote data that lies 1.5*IQR away from the ends of the box, where IQR (the interquartile range) is calculated as the distance between the 25th and 75th percentiles. The boxplots in Figure 9 indicate that the spread seems comparable among the different disconnects and calibrations for the attenuator's $|S_{21}|$ measurements. Residual analysis for the attenuator's $\text{Arg}\{S_{21}\}$ measurements showed similar results.

IV. Conclusions

We described a two-stage nested design that allows us to quantify random uncertainties for repeated S-parameters measurements, and focused on variations due to multiple calibrations, disconnects, and repeat measurements. Furthermore, we presented results for a series of coaxial measurements performed by making use of OSLT calibrations with Type-N connectors, and discovered our variance estimates did not gradually increase with frequency, but rather rose and fell unexpectedly within the measured frequency range. Neither the estimated calibration or disconnect variances consistently dominated the overall variance throughout the measured frequency range. And in the case presented, the systematic uncertainty values were significantly greater than the random uncertainties values at higher frequencies.

The method we presented may be used for examining a host of other calibration techniques with varying connector sizes and environments, such as waveguide and on-wafer. We conjecture other such combinations could provide dramatically different results.

From our experience with this experiment, because the variation due to error is negligible compared to the variation due to disconnect and calibration, we conclude that random uncertainties can best be captured by making multiple disconnects of any devices-under-test of interest as well as performing multiple calibrations. Simply making repeat measurements on a device without disconnecting and reconnecting it offers little insight regarding the overall random uncertainty.

ACKNOWLEDGEMENTS

*This work was supported by the U.S. government, and is not subject to U.S. copyright.

The authors thank Kevin Coakley, Mike Frey, Paul Hale, Michael Janezic, Aric Sanders, Jolene Splett, Sarah Streett, and Mitch Wallis for their helpful comments.

REFERENCES

[1] J. A. Jargon, C. H. Cho, D. F. Williams, and P. D. Hale, "Physical Models for 2.4 mm and 3.5 mm Coaxial VNA Calibration Kits Developed within the NIST Microwave Uncertainty Framework," *85th ARFTG Microwave Measurement Conference*, Phoenix, AZ, May 2015.

[2] J. A. Jargon, D. F. Williams, and P. D. Hale, "Developing Models for Type-N Coaxial VNA Calibration Kits within the NIST Microwave Uncertainty Framework," *87th ARFTG Microwave Measurement Conference*, San Francisco, CA, May 2016.

[3] F. A. Graybill, *Theory and Application of the Linear Model.* North Scituate, MA: Duxbury Press, 1976.

[4] D. C. Montgomery, *Design and Analysis of Experiments*, 8th ed. New York: Wiley, 2013.

[5] S. R. Searle, *Linear Models*. New York: Wiley, 1971.

Koepke, Amanda; Jargon, Jeffrey.
"Quantifying Variance Components for Repeated Scattering-Parameter Measurements."
Paper presented at 90th ARFTG Microwave Measurement Symposium, Boulder, CO, United States. November 30, 2017 - December 1, 2017.

AC-DC Difference of a Thermal Transfer Standard Measured with a Pulse-Driven Josephson Voltage Standard

G. Granger[*], N. E. Flowers-Jacobs[†], and A. Rüfenacht[†]

[*] National Research Council Canada, Ottawa, Ontario, K1A 0R6, Canada
ghislain.granger@nrc-cnrc.gc.ca

[†] National Institute of Standards and Technology, Boulder, CO 80305, USA

***Abstract* — An upgraded pulse-driven AC Josephson voltage standard system at the National Research Council Canada allows the generation of quantum-accurate root-mean-square voltages up to 1 V at frequencies down to 10 Hz. The system is used to determine the AC-DC difference of a commercial thermal transfer standard, and an uncertainty budget is presented. Such quantum voltage systems could replace aging multi-junction thermal converters.**

***Index Terms* —Josephson arbitrary waveform synthesizers, Josephson voltage standards, pulse-driven thermal transfer standards, thermal converters.**

I. Introduction

Traditionally, AC voltage has been traceable to DC voltage by means of the AC-DC transfer difference of sets of thermal voltage converters (TVCs). Joule heating at the resistive element inside a TVC gives rise to a local temperature increase, which generates an electromotive force across an output thermocouple. The AC-DC difference is defined as the relative difference between the AC voltage and the polarity-averaged DC voltages that produce the same output [1].

Acquired and setup in 2008, the pulse-driven AC Josephson voltage standard (also known as the Josephson Arbitrary Waveform Synthesizer [JAWS]) at the National Research Council Canada (NRC) generated voltages up to 250 mV at frequencies ranging from 2.5 kHz to 1 MHz. It was used in international comparisons of a commercial thermal transfer standard (TTS) [2,3]. Recent advances in JAWS technology resulted in higher voltages and lower frequencies [4-6].

In this paper, we report on measurements performed at NRC with an upgraded JAWS system, with new components acquired from the National Institute of Standards and Technology (NIST)[1] and High Speed Circuit Consultants (HSCC)[2] to generate waveforms with root-mean-square (rms) voltages as high as 1 V and frequencies as low as 10 Hz. We compare the AC-DC voltage difference of a commercial TTS at 0.6 V obtained from the JAWS to results obtained by conventional methods. A basic lead correction scheme is applied to the data taken at higher frequencies, and a quantum locking range study and an uncertainty budget are presented.

[1] Contribution of the U.S. Government, not subject to copyright.

[2] Certain commercial equipment, instruments, or materials are identified to facilitate understanding. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment are necessarily the best available for the purpose.

II. Upgraded JAWS System at NRC

The upgraded JAWS system at NRC is based on a NIST chip with 4 subarrays of 12,810 Josephson junctions (JJs) each. Biases are provided by the HSCC ABG-2 bitstream generator and AWG-1 low-frequency compensation source. The JAWS chip is mounted in a cryoprobe dipped in liquid helium. The 28.8-gigabit-per-second bitstream generator combines a digital pulse pattern (i.e., the 3-level, sigma-delta, digital-to-analog conversion of the desired waveform) with a 14.4 GHz continuous-wave microwave signal. The resulting train of current pulses is injected into the arrays through on-chip Wilkinson dividers [6]. The ABG-2 is connected to the NRC 10 MHz reference frequency signal to ensure frequency traceability. DC blocks inside the cryoprobe effectively serve as a high-pass filter for the input pulse train. The low-frequency current components are restored using the battery-operated AWG-1.

Current margins are determined using a digitizer and a current bias sweep from additional AWG-1 channels. If the current pulses are within the margins of the $n = 1$ quantized steps of the JJs, the time integral of each resulting voltage pulse is precisely quantized to a flux quantum, and the fundamental component of the output voltage pulse train becomes calculable. Due to the limited bandwidth of the cryoprobe, the high-frequency components of the output voltage pulse train are filtered, such that the desired quantum-accurate waveform remains.

III. AC-DC Difference of the TTS

We first determine the AC-DC difference of the TTS at 0.6 V in its 700 mV range (input resistance ~10 MΩ) using conventional references: a multi-junction thermal converter (MJTC) (0.01 kHz-50 kHz) and another TTS (100 kHz-500 kHz).

To compare the TTS to the JAWS system, the JAWS output is connected to the TTS input with a short adapter directly at the top of the cryoprobe to minimize losses in AC voltage that result from the 1.4 m long leads between the chip and the TTS. AC-DC measurements are made by generating a sequence of DC+, DC-, and AC voltages of calculable amplitudes with the JAWS. The results are shown in Fig. 1:

Flowers-Jacobs, Nathan; Rufenacht, Alain; Granger, Ghislain.
"AC-DC Difference of a Thermal Transfer Standard Measured with a Pulse-driven Josephson Voltage Standard."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

there is excellent agreement: the difference between the results from two methods is zero within uncertainties.

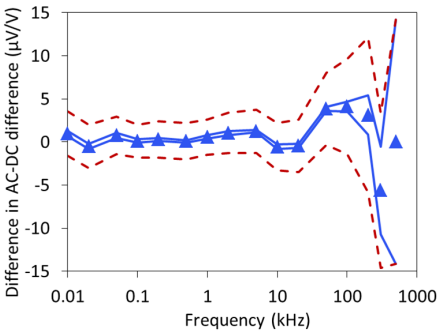


Fig. 1. Difference between the AC-DC difference of the TTS at 0.6 V in its 700 mV range measured using the JAWS and conventional references (triangles). Expanded uncertainties at $k = 2$ are shown for the JAWS (solid line) and the conventional references (dashed line).

IV. Uncertainty Analysis

A typical standard deviation of the mean of 10 sequences of AC-DC measurements is used for the type A uncertainty. To appreciate whether the AC-DC difference depends on the bias parameters giving optimal margins, we repeat the AC-DC measurements at 1 kHz with different dithers in the bias parameters. In the case shown in Fig. 2, the applied dither ranges from −5% to 5% of the bias parameter settings. A linear fit and its statistics reveal an uncertainty on the intercept at 0 % dither of 0.027 μV/V.

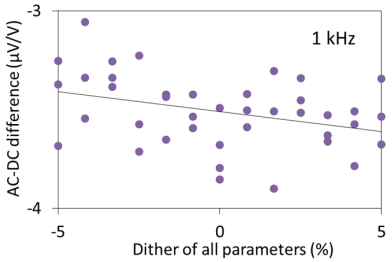


Fig. 2. Measured quantum locking range and linear fit showing the AC-DC difference obtained using the JAWS at 0.6 V and 1 kHz as a function of percentage of dither of all bias parameters, including the amplitude and phase of the microwave channels, the mixer amplitude, and the low-frequency bias currents.

To facilitate the margins optimization for the DC patterns, a 1 kHz tone with an amplitude 80 dB below the DC signal is added. The output voltage is directly proportional to the TTS rms input voltage; therefore, assuming a uniform distribution, the uncertainty component due to this added tone is 0.0029 μV/V. We also take into account the uncertainty from the frequency signal, estimated at 0.042 nV/V. Small AC and DC pattern offsets originating mostly from the finite number of Josephson junctions are taken into account based on the resolution of the offset correction applied to get the AC-DC difference. This uncertainty is 0.029 nV/V. For a given cable at frequencies >10 kHz, lead corrections are empirically quadratic in frequency and set by assuming perfect agreement at 500 kHz. The uncertainty for the lead corrections is obtained by scaling the uncertainty from the comparison to the other TTS at 500 kHz. The uncertainty budget for the data taken at 0.6 V is shown in Table 1. It combines to an expanded uncertainty ($k = 2$) of 0.27 μV/V up to 10 kHz.

Table 1. Uncertainty budget for the AC-DC difference of the TTS at 0.6 V in the 700 mV range measured against the JAWS. f is the frequency in hertz.

Standard uncertainty component (μV/V)					
Type A	Dither	1 kHz tone	Freq. offset	Pattern offsets	Lead Correction ($f > 10^4$ Hz)
0.13	0.027	0.0029	4.2×10^{-5}	2.9×10^{-5}	$2.9\times10^{-11} f^2$

V. Conclusion

We measured the AC-DC difference of a commercial thermal transfer standard using the JAWS at 0.6 V and found excellent agreement with the results obtained using conventional references. An uncertainty budget for the method involving the JAWS achieved a combined expanded uncertainty ($k = 2$) of 0.27 μV/V at 10 kHz and below. Lead correction uncertainties have been estimated for frequencies of 20 kHz–500 kHz. We expect the JAWS to replace aging thermal converters as the primary source of AC-DC difference traceability in the near future.

Acknowledgement

We thank Piotr Filipski, Sam Benz, and Steve Waltman for sharing their expertise and Carlos Sanchez for his guidance.

References

[1] B. D. Inglis, "Standards for AC-DC Transfer," *Metrologia*, vol. 92, pp. 191 – 199, 1992.

[2] T. E. Lipe *et al.*, "An International Intercomparison of Quantum-Based AC Voltage Standards," 2012 *IEEE International Instrumentation and Measurement Technology Conference Proceedings*, Graz, pp. 98-102, 2012.

[3] P. S. Filipski *et al.*, "International comparison of quantum AC voltage standardsfor frequencies up to 100 kHz," *Measurement*, vol. 4, pp. 2218–2225, 2012.

[4] S. P. Benz *et al.*, "One-Volt Josephson Arbitrary Waveform Synthesizer," *IEEE Trans Appl. Supercond.*, vol. 25 (no. 1), 1300108, February 2015.

[5] O. F. Kieler *et al.*, "Towards a 1 V Josephson Arbitrary Waveform Synthesizer", *IEEE Trans. Appl. Supercond.*, vol. 25 (no. 3), 1400305, June 2015.

[6] N. E. Flowers-Jacobs *et al.*, "Two-Volt Josephson Arbitrary Waveform Synthesizer Using Wilkinson Dividers," *IEEE Trans. Appl. Supercond.*, vol 26 (no. 6), 1400207, September 2016.

Flowers-Jacobs, Nathan; Rufenacht, Alain; Granger, Ghislain.
"AC-DC Difference of a Thermal Transfer Standard Measured with a Pulse-driven Josephson Voltage Standard."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Metrological and legal traceability of time signals

Demetrios Matsakis[1], Judah Levine[2], and Michael A. Lombardi[2]
[1]*United States Naval Observatory, Washington, DC, USA*
[2]*Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado, USA*

ABSTRACT

Metrological traceability requires an unbroken chain of calibrations that relate to a reference, with each calibration having a documented measurement uncertainty. In the field of time and frequency metrology, the desired reference is usually Coordinated Universal Time (UTC), or one or more of its official realizations, termed UTC(*k*), and traceability to UTC is a legal requirement for many entities. Traceability to UTC can be established in three areas – frequency, time interval, and time-of-day synchronization, but this paper focuses solely on the traceability of time signals used for synchronization. We first examine the definition of traceability, then discuss how traceability can be established via the reception of time signals transmitted by satellites and network time servers, followed by a discussion of how these signals can meet the synchronization and traceability requirements of the financial and electric power industries. Not all of the available UTC time signals are considered in this paper, as we primarily focus on direct broadcast and common-view Global Positioning System (GPS) signals, with uncertainties measured in nanoseconds, and Network Time Protocol (NTP) signals, with uncertainties measured in microseconds and milliseconds.

I. INTRODUCTION

The International Vocabulary of Metrology (VIM) defines metrological traceability in section 2.41 (6.10) as "the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty" [1]. The VIM is endorsed by the International Bureau of Weights and Measures (BIPM), the International Electrotechnical Commission (IEC), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Organization for Standardization (ISO), the International Union of Pure and Applied Chemistry (IUPAC), the International Union of Pure and Applied Physics (IUPAP), the International Organization of Legal Metrology (OIML), and the International Laboratory Accreditation Cooperation (ILAC). Of these ILAC has a more detailed definition, which includes traceability to an international or national measurement standard, a documented measurement uncertainty, a documented measurement procedure, accredited technical competence, and comparisons to the international system of units (SI) with calibrations at regular intervals. The International Telecommunications Union (ITU) adopted the following definition in 2013: "the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties" [2]. The ITU definition is essentially the same definition found in previous editions of the VIM. The most noteworthy change in the current VIM definition is probably the use of the words "documented" and "calibrations" as opposed to "comparisons" in the earlier definition, but the meaning and intent remain the same.

Determining whether enough evidence exists to establish traceability is usually the role of an auditor or assessor who visits the laboratory or facility where the measurements are performed. These auditors or assessors are usually working on behalf of an accreditation body such as ILAC and/or a standards organization such as ISO. In some cases, however, the determination of whether measurements are traceable is made by a regulatory agency, such as, in the case of financial markets, the U. S. Securities & Exchange Commission (SEC). In cases where losses or damages are incurred, the final determination of whether traceability was properly established may be made in a court of law. Those charged with the responsibility of proving or disproving traceability often refer to the internationally accepted definition of metrological traceability provided in the VIM [1]. This definition consists of four key parts, as described and expanded upon in the following sections.

Part 1 - *Traceability is "the property of a measurement result"* **-** This means that the concept of traceability only applies to measurement results and not to other things. Therefore:

- Traceability is not a property of a system. For example, traceability is not a property of the Global Positioning System (GPS) but a traceable measurement that involves GPS can be made.

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

- Traceability is not a property of an instrument. For example, traceability is not a property of a time interval counter or even a cesium clock but a traceable measurement that involves these instruments can be made.

- Traceability is not a property of an organization or laboratory. For example, simply being the United States Naval Observatory (USNO) or the National Institute of Standards and Technology (NIST) does not guarantee that all measurements made at USNO or NIST are traceable, nor does it guarantee that all measurements referenced to USNO or NIST signals are traceable.

Part 2 – *A traceable measurement "can be related to a reference"* **–** For nearly all areas of metrology, including time and frequency, the ultimate measurement reference is the International System (SI) of units. The SI units are definitions of ideal values and as such have zero uncertainty. However, they are not physical standards, and establishing traceability requires a real measurement that involves a comparison against a physical standard.

Only two SI units apply to time and frequency, the second (s) and the hertz (Hz). The second is the standard unit for time interval, and one of the seven base units of the SI. Since 1967, it has been defined as "The duration of 9,192,631,770 periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium-133 atom. [3]." The hertz is the standard unit for frequency. It represents events per second (the events are usually pulses or cycles in an electrical signal), and is defined as s^{-1} in SI parlance. The hertz is one of 21 named SI units that are derived from the base units.

The world's best approximation of the SI units for time and frequency, and thus the ultimate reference for establishing traceability, is Coordinated Universal Time (UTC), an atomic time scale based on the SI definition of the second. UTC is computed by the BIPM in France by performing a weighted average of data collected from local time scales located at more than 70 timing laboratories [4]. UTC itself exists only on paper, and is defined through its difference with the local time scales, known as UTC(*k*), of participating laboratories. It is published monthly in the BIPM *Circular T* (Figure 1). The *Circular T* is typically published around the tenth day of a month, and covers the previous month. Thus, the latency of the published measurement results typically ranges from 10 to 45 days.

CIRCULAR T 357 ISSN 1143-1393

2017 OCTOBER 10, 13h UTC

BUREAU INTERNATIONAL DES POIDS ET MESURES
ORGANISATION INTERGOUVERNEMENTALE DE LA CONVENTION DU METRE
PAVILLON DE BRETEUIL F-92312 SEVRES CEDEX TEL. +33 1 45 07 70 70 FAX. +33 1 45 34 20 21 tai@bipm.org

The contents of the sections of BIPM *Circular T* are fully described in the document " Explanatory supplement to BIPM Circular T " available at ftp://ftp2.bipm.org/pub/tai/publication/notes/explanatory_supplement_v0.1.pdf

1 - Difference between UTC and its local realizations UTC(k) and corresponding uncertainties. From 2017 January 1, 0h UTC, *TAI-UTC* = 37 s.

Date 2017 0h UTC			AUG 29	SEP 3	SEP 8	SEP 13	SEP 18	SEP 23	SEP 28	Uncertainty/ns			Notes
		MJD	57994	57999	58004	58009	58014	58019	58024	u_A	u_B	u	
Laboratory *k*						[*UTC-UTC(k)*]/ns							
AOS	(Borowiec)		-2.6	-3.6	-4.5	-4.3	-4.0	-4.2	-4.1	0.6	3.0	3.0	
APL	(Laurel)		1.6	-2.9	-2.2	-1.8	-2.8	-2.8	-2.7	0.4	10.9	10.9	
AUS	(Sydney)		436.8	428.0	411.4	413.3	413.1	405.2	387.6	0.4	5.9	5.9	
BEV	(Wien)		36.2	36.8	36.9	22.7	24.4	22.3	11.1	0.4	2.8	2.8	
BIM	(Sofiya)		6864.7	6893.9	6941.7	6995.9	7064.0	7121.9	7137.2	0.7	3.0	3.1	
BIRM	(Beijing)		37.7	33.5	28.6	31.8	34.3	33.3	30.3	0.7	2.8	2.9	
BOM	(Skopje)		-839.4	-837.9	-834.4	-836.4	-833.0	-841.0	-842.1	0.4	3.0	3.0	
BY	(Minsk)		-0.6	-0.5	-0.9	-1.4	3.0	6.6	3.6	1.5	9.3	9.4	
CAO	(Cagliari)		-	-	-	-	-	-	-				
CH	(Bern-Wabern)		27.9	25.9	22.3	21.7	21.3	23.4	27.9	0.4	1.9	1.9	
CNES	(Toulouse)		27.3	28.6	31.8	32.5	27.9	25.1	23.2	0.4	4.3	4.3	
CNM	(Queretaro)		-1.1	-4.8	-1.1	3.1	-4.4	-0.1	-0.4	2.5	11.1	11.4	
CNMP	(Panama)		-41.5	-35.5	-22.7	15.2	24.4	23.9	21.3	0.6	7.2	7.2	
DFNT	(Tunis)		19872.3	20054.3	20250.8	20458.3	20650.9	20834.1	21007.3	0.7	20.0	20.1	
DLR	(Oberpfaffenhofen)		44.4	43.8	45.2	53.7	69.2	238.6	285.0	0.7	3.0	3.1	(1)
DMDM	(Belgrade)		-0.8	-2.9	3.9	13.0	7.0	14.3	13.8	0.4	2.8	2.8	

Figure 1. A portion of Section 1 of the BIPM's monthly *Circular T* document.

The UTC(*k*) are the outputs of physical standards that continuously realize the SI units of time and frequency, and can thus serve as the reference for real measurements. Section 1 of the *Circular T* shows UTC – UTC(*k*) for every contributing laboratory at 5-day intervals, allowing each participant to establish traceability to the SI. Three additional points should be noted about UTC:

- The measurements of the UTC(k) time scales never stop and thus traceability can be continuously established, not just at irregular intervals as is the case in other fields of metrology. These measurements constitute BIPM key comparison CCTF-K001.UTC which has been ongoing since 1977.

- UTC is the ultimate reference not only for frequency and time interval, but also for everyday time-of-day synchronization.

- UTC has been synchronized, since its inception, to stay within 0.9 seconds of the predicted value of UT1, the rotation angle of the Earth and effectively the successor to the no-longer existent Greenwich Mean Time. This synchronization involves the aperiodic insertion of leap seconds [5].

Part 3 - *Traceability requires "a documented unbroken chain of calibrations"* – This part of the definition tells us that claims of traceability must always be supported by actual measurements. It is incorrect, for example, to say that because a signal originates from NIST or USNO that it is therefore traceable without calibration. A calibration is a comparison between a reference and a device under test (DUT) that is conducted by collecting and analyzing measurements. In addition:

- The chain of calibrations must be documented. The amount of documentation will depend on the requirements of the organization or sector that needs to provide proof of traceability. In the United States, audit trail standards for the financial community are set using Rule 613 [6]. This documentation should cover the entire traceability chain; from the SI to the end-user.

- There is no restriction or limit on the number of calibration steps (or links in the traceability chain), but a simple traceability chain is easier to maintain and/or to demonstrate to auditors or assessors. The unbroken chain of calibrations must trace back from the measurement in question to a representation of the SI, which means that it will normally extend from the end user to a UTC(k) reference maintained by a laboratory such as NIST or the USNO.

Part 4 - *Each calibration in a traceability chain is "contributing to the measurement uncertainty"* – Traceability cannot exist without knowledge of the measurement uncertainty. Measurement uncertainty is defined in the VIM as "The parameter, associated with the result of a measurement, that characterizes the dispersion of values that could reasonably be attributed to the measurand." The measurand is the "particular quantity subject to measurement," [1] which in our case can be either time referenced to a UTC realization, time interval, or frequency.

Determining the uncertainty of a measurement result involves knowing the uncertainty of each calibration, or every link in the traceability chain, between UTC and the end user. The uncertainty of some links may be negligible and easy to document, for example the *Circular T* provides the uncertainty between UTC and UTC(k), but determining the uncertainty of the links that connect UTC(k) to the end user is often the most difficult part of establishing traceability. It requires knowledge of the uncertainty of the path delay between the time signal source and the user, as well as knowledge of the uncertainty of the calibration of the equipment used to receive the time signal, which could be, for example, a GPS receiver or a computer with network time protocol (NTP) client software.

The internationally recognized standards document for measurement uncertainty analysis is called the "Guide to expression of uncertainty in measurement," colloquially known as the GUM [7]. The GUM recommends that:

- Every parameter that contributes to the measurement uncertainty should be evaluated with either the Type A or the Type B method. Type A parameters are evaluated by the statistical analysis of a series of measurements, for example by use of the standard deviation, Allan deviation, or a similar statistic. Type B parameters are evaluated by non-statistical means. An example might be a single measurement of a cable delay that is applied as a constant in all subsequent uncertainty analysis.

- It is customary to combine the Type A and Type B parameters by taking the square root of the sum of the squares, and then multiplying the result by the coverage factor, k. The coverage factor relates to the range of the distribution and reflects the probability (which is approximately 68.3% for $k = 1$ or 95.5% for $k = 2$), that a given measurement result will fall within this range. This method for estimating the combined uncertainty is utilized on the *Circular T* (Figure 1) where $k = 1$ and all uncertainties are stated in units of nanoseconds. The Type A uncertainty, u_A, the Type B uncertainty, u_B, and the combined uncertainty, u, are listed in separate columns. The equation for determining u is

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

$$u = k\sqrt{{u_A}^2 + {u_B}^2}\,. \qquad (1)$$

II. OFFICIAL TIME IN THE UNITED STATES

As discussed in Section I, international timekeeping is organized by a system in which individual timing laboratories can realize UTC in real time, and those realizations are termed UTC(*k*). The BIPM publishes its monthly *Circular T* which shows the time difference between UTC and its local realizations in the form of UTC – UTC(*k*).

Most nations have only one laboratory listed on the *Circular T*, but the United States has four. In addition to the USNO and NIST, the Naval Research Laboratory (NRL) and the Applied Physics Laboratory (APL) at John Hopkins University also contribute to UTC and, in theory, metrological traceability could also be established through either NRL or APL. However, the local UTC(*k*) time scales maintained by NRL and APL exist primarily for internal research purposes. Only the USNO and NIST distribute time and frequency signals in the United States, and jointly hold the responsibility of being the official U. S. timekeeper.

The America COMPETES Act of 2007 [8] specifies that the official time is UTC, as "interpreted or modified by the Secretary of Commerce, in coordination with the Secretary of the Navy." Ironically, until this bill was passed in 2007, official U. S. time was "mean solar time" as defined in the Calder Act of 1918. In practice, this means NIST (an agency of the U. S. Department of Commerce) is officially responsible for determining the UTC that is used for the financial and electric power sectors, while the USNO is the source of UTC for the Department of Defense and GPS, as specified in DoD Instruction 4650.06 [9]. Therefore, the legal traceability path for time measurements in the United States must involve either NIST or the USNO although, as noted above, establishing traceability to any UTC(*k*) can be used to establish metrological traceability to all of them, at the expense of increased complexity.

III. ESTABLISHING TRACEABILITY VIA THE DIRECT RECEPTION OF GPS TIME SIGNALS

The Global Positioning System (GPS) has its own time scale, known as GPS time, but the satellites broadcast parameters in subframe 4, page 18, of the GPS navigation message that receivers can apply to convert GPS time to a prediction of UTC(USNO). Although GPS time could in principle be used for traceability if the user were to carefully apply a considerable number of corrections, it is intended only for positioning and does not include leap seconds. It should not be used by time users, and nearly all GPS receivers apply the UTC corrections to their time determination by default. It is not even possible to disable the corrections for many receiver models. A receiver can obtain the UTC correction parameters from any satellite, but should use the satellite whose information was most recently refreshed. The UTC offset correction, Δt_{UTC}, is computed as [10]

$$\Delta t_{\mathrm{UTC}} = \Delta t_{\mathrm{LS}} + A_0 + A_1\,(t_{\mathrm{E}} - t_{\mathrm{ot}} + 604800(WN - WN_{\mathrm{t}})), \qquad (2)$$

where

Δt_{LS} is the number of leap seconds introduced into UTC since GPS time began on January 6, 1980,

A_0 is the constant UTC offset parameter expressed in seconds,

A_1 is a dimensionless frequency offset value that allows the correction of the time error accumulated since the UTC reference time, t_{ot}, which is when A_0 was last determined,

t_{E} is GPS time (also known as the time of interest or the time being converted to UTC),

604800 is a constant that equals the number of seconds in one week.

t_{ot} is the reference time for UTC data,

WN is the GPS week number, and

WN_{t} is the UTC reference week number.

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

The first part of Eq. (2), $\Delta t_{LS} + A_0$, takes care of most of the UTC correction. The Δt_{LS} term is the large, integer second part of the correction, equal to the number of leap seconds that have occurred since the start of the GPS time scale. The A_0 term is the small, nanosecond part of the correction, equal to the difference between the GPS and UTC(USNO) second markers. It is broadcast in units of seconds, but is typically $< 1 \times 10^{-8}$ s, or < 10 ns in magnitude. The second part of Eq. (2) fine tunes the UTC output of a GPS clock by applying a dimensionless frequency offset, provided by A_1, as a drift correction for the interval between the time specified by t_{ot} and WN_t and the current time. This is normally a sub-nanosecond correction, because A_0 is normally updated in the GPS broadcast more than once per day and the drift correction supplied by A_1 is typically near 1 ns per day.

The UTC(USNO) prediction is based upon an extrapolation of the observed difference between GPS and UTC(USNO) from the start to the end of the previous day. The accuracy of that prediction, for the signal in space, is less than 1 ns in practice. To illustrate this, Figure 2 shows the differences between GPS time and UTC(USNO) modulo 1 s (to remove the leap second differential), as well as the difference between GPS predicted UTC and UTC(USNO) for the period from 2013 to the summer of 2017. More recent data can be obtained from the http://tycho.usno.navy.mil and ftp://tycho.usno.navy.mil/pub/gps. The procedures that GPS uses to deliver time are documented in the Interface Control Documents ICD202 and ICD200 [10, 11], where the official accuracy is currently (and very conservatively) listed as 90 ns. The actual data show that if we can accept a latency of up to two days, the time obtained from the GPS signal in space as transmitted by the satellite can be considered directly traceable to UTC(USNO), with an uncertainty of a few nanoseconds or less. Alternately, the real-time signal-in-space broadcast of UTC(USNO) can be considered to have negligible latency but with an additional uncertainty component of order 1 ns.

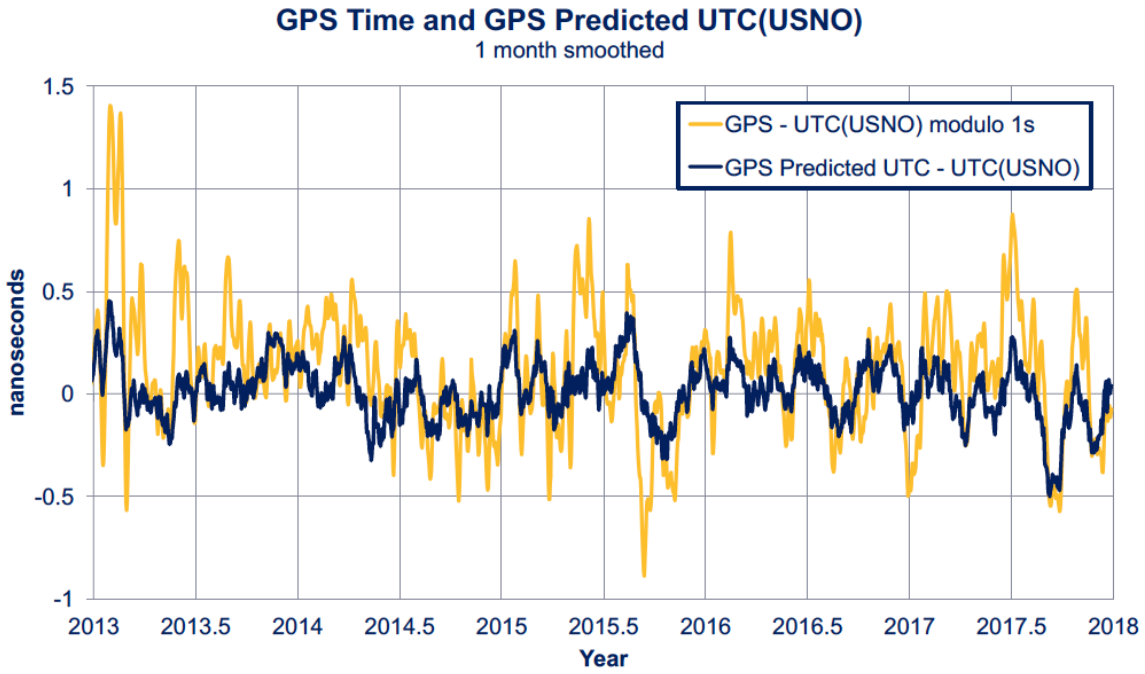


Figure 2. Time differences between GPS timing signals and UTC(USNO), from 2013 to 2017 as received at the USNO, courtesy Stephen Mitchell (USNO)

Section IV of the BIPM's *Circular T* document also publishes values for UTC – UTC(USNO via GPS) as well as UTC – UTC(SU via GLONASS), but with larger uncertainties. The *Circular T* does not currently include anything similar for the European Galileo satellites because Galileo's time is based on an average of UTC realizations. However, as described in the next section, a legal traceability chain that complies with the VIM definition could also be established by use of Global Navigation Satellite System (GNSS) signals as with GPS.

The uncertainty of the GPS signal in space with respect to UTC is ~1 ns, as shown in Figure 2, but will be considerably larger when received on Earth. This is due to many factors, including the quality and calibration of the user's receiving equipment,

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

errors in the calibration of the user's antenna and antenna cable, antenna coordinate errors, environmental effects, ionospheric and tropospheric delays, multipath signal reflections, and other factors. For these reasons, the synchronization uncertainty of a GPS disciplined clocks ($k = 2$) will be ~10 ns in the best case and ~1 µs in the worst case. It is the responsibility of the user to provide enough documentation to support and if necessary, to defend, the uncertainty that they claim. Detailed methods for evaluating the uncertainty of GPS disciplined oscillators and clocks are provided in [12].

IV. ESTABLISHING TRACEABILITY VIA COMMON-VIEW GPS TIME SIGNALS

NIST and other national timing laboratories now distribute their UTC time scales to their customers by publishing the difference between UTC(k) and the prediction of UTC(USNO via GPS) for every satellite. These data can be used to compute the difference between a user's local clock and UTC(k) by observing the same satellite at the same time and subtracting the two measurements. However, the user is responsible for the calibration of their common-view equipment and needs also to consider the latency of the published data, which could be hours or days. A similar use could be made of any GNSS signal, provided that the reference laboratory makes the common-view information available.

When a common-view comparison is conducted with a UTC(k) laboratory (Figure 3), a reference system produces continuous measurements of *UTC(k) – GPS*. A system at the customer's site simultaneously produces continuous measurements of *Local Clock – GPS*. The measurements from the reference system and the customer's system are subtracted from each other, the time broadcast by GPS falls out of the equation, and the result is an estimate of *UTC(k) – Local Clock*. The GPS signals in this case are simply vehicles used to transfer or relay time from UTC(k) to the customer's site. The common-view equation for this example is

$$(UTC(k) - GPS) - (Local\ Clock - GPS) = (UTC(k) - Local\ Clock) + (e_{SA} - e_{SB}), \qquad (3)$$

where the components that make up the ($e_{SA} - e_{SB}$) error term include delay differences between the two sites caused by ionospheric and tropospheric delays, multipath signal reflections, environmental conditions, or errors in the GPS antenna coordinates. These factors are either measured or estimated and applied as a correction to the measurement of the local clock, with the uncorrected portions contributing to the uncertainty.

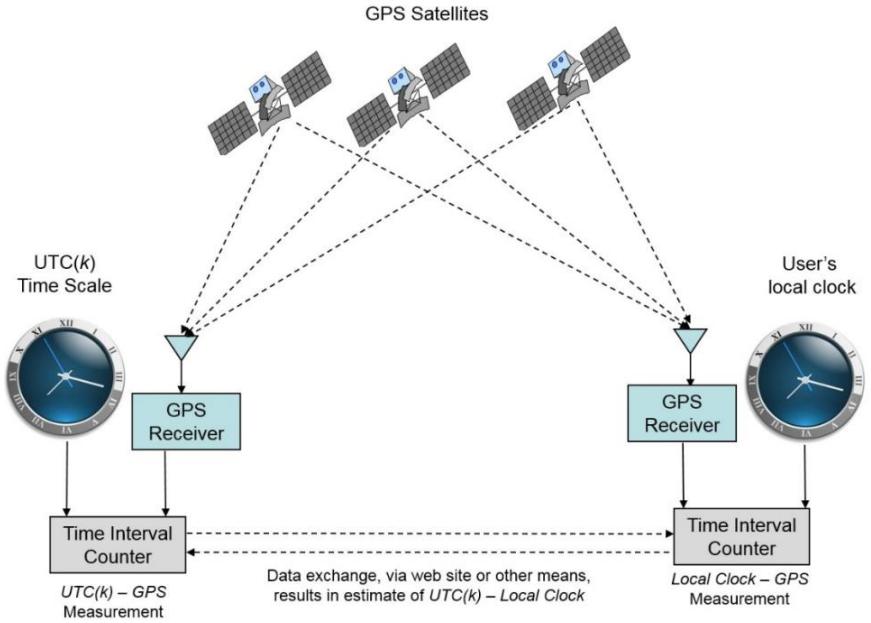


Figure 3. A common-view GPS comparison with a UTC(k) laboratory.

NIST and other national timing laboratories also offer formal common-view and all-in-view (described below) based services, in which each customer receives a common-view system whose receiver, antenna, and cable delays have all been calibrated

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

prior to shipment. In the case of the NIST services, both NIST and the customer simultaneously upload their measurements to a data server every 10 minutes, the data processing is automated, and the results are published in near real-time. NIST also offers a disciplined clock (NISTDC) service (Figure 4) where the common-view measurement results are used to lock the customer's clock to UTC(NIST), just as a GPS disciplined clock is locked to UTC(USNO via GPS). The uncertainty of the NIST common-view services with respect to UTC(NIST) is reported monthly to each customer and is typically near 10 ns (k = 2) [13].

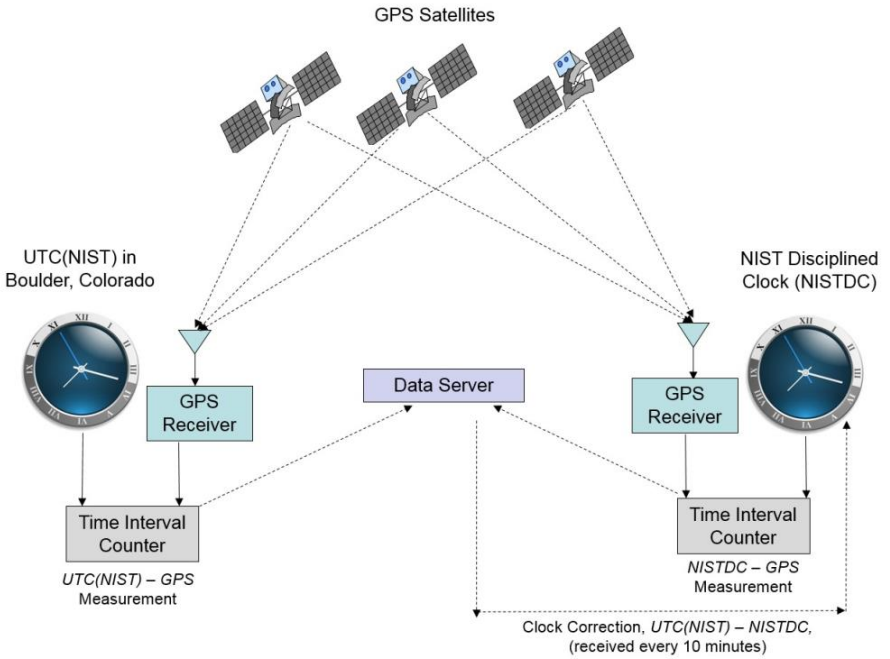


Figure 4. NIST disciplined clock system utilizing the common-view GPS method.

A measurement technique similar to common-view is known as all-in-view. For all-in-view, the time difference between the laboratory clock and all satellites in view is first averaged, and the timing difference is estimated from the difference between laboratory averages. This method has been adopted by the BIPM because it affords better signal to noise over long distances, when there may not be many satellites visible at both labs at the same time. Verifying all-in-view traceability is more complicated than in the common-view method because it requires adding a component to the error budget that accounts for the difference between the time as broadcast by the individual satellites. This is because errors in the broadcast ephemerides of the satellites, which are attenuated in common-view, must be allowed for and because the method depends on the availability of precise ephemerides and clock products. While these error sources can be estimated from past data, and measured after-the-fact, all-in-view would also remain sensitive to the same local station-errors that affect common-view.

V. THE TRACEABILITY OF NETWORK TIME PROTOCOL (NTP) SIGNALS

As discussed in Sections III and IV, GPS and other GNSS signals can provide traceability with uncertainties measured in nanoseconds. Another important and widely used source of traceable time signals are the NTP time servers that reside on the public Internet, albeit with much higher uncertainties measured in microseconds or milliseconds. The primary purpose of NTP is the synchronization of computer clocks and network appliances. The demand for these services is so high that many billions of NTP synchronization requests are currently received every day [14]. At this writing (November 2017) the average number of NTP requests received per second is about 460 000 at NIST and about 15 000 at the USNO.

The NTP services measure the round trip delay between the server and the client, assume that the one-way delay from the server to the client is equal to half of the round trip delay, and then correct the time received by the client by the one-way delay. This correction compensates for most, but not all, of the propagation delay between the server and the client. The standard NTP equation is

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

$$TD = \frac{((T_2 - T_1) + (T_3 - T_4))}{2}. \quad (4)$$

where *TD* is the time difference between the server and client clocks, T_1 is the time when the client made the request, T_2 is when the request was received by the server, T_3 is when the server transmitted its response, and T_4 is when the time packets transmitted by the server arrive at the client. Using these same four time stamps, the round trip delay between the client and server [15] is computed by the client as

$$RT_{Delay} = (T_4 - T_1) - (T_3 - T_2). \quad (5)$$

The division by two in Eq. (4) assumes that the delay from the server to the client is equal to one half of the round trip delay. If this assumption were true, the network would be symmetric, meaning that the path delays to and from the server would be equal and that dividing by two would compensate for all delays. In practice, however, networks are asymmetric. This means that the incoming and outgoing delays are not equal and that the difference in delays will contribute uncertainty to the time received by the client. The worst-case time uncertainty that can be added by network asymmetry is 50% of the round trip delay [15, 16], a situation that could, of course, only occur if 100% of the delay were in one direction.

An additional assumption in the use of Eqs. (4) and (5) is that the four time values are measured with negligible uncertainty. This assumption is questionable when the Network Time Protocol application runs as a normal user process on a system that supports a graphical user interface. The overhead in the communication between the user process and the clock of the operating system in this case may not be negligible, especially on a heavily-loaded system or one that is supporting web-based services. In addition, all of the time values are large quantities, and the potential loss of significance when the difference between two large quantities is very small is an important detail in the implementation of the software.

If the servers have been properly synchronized to a UTC(*k*) source such as UTC(NIST) or UTC(USNO), network asymmetry will probably dominate the uncertainty of NTP time transfer on a wide-area network. Because the maximum uncertainty cannot exceed 50% of the round trip delay, the uncertainty of NTP is likely to be much smaller on a local area network (LAN) than it is on a wide area network (WAN) such as the public Internet. On a WAN, care should be taken not to underestimate the uncertainty when establishing a traceability chain via NTP. A study of the USNO NTP service involving multiple servers and clients revealed semi-persistent systematic time errors as large as 100 ms, often due to the rerouting of packets and network congestion [16]. Even when Internet conditions are favorable, it is uncommon to have a network asymmetry smaller than a few percent of the round trip delay. Thus, if the round trip delay is 100 ms, it is reasonable to assume that the uncertainty is not less than a few milliseconds [17]. On a controlled LAN, uncertainties much smaller than 100 μs have been reported, but the remaining uncertainties may be difficult to characterize, including asymmetric delays in network interface cards and client and server instability [18]. To properly estimate the uncertainty of an NTP link, the received packets should be compared to a UTC(*k*) time source, such as UTC(NIST), UTC(USNO), or UTC(USNO via GPS). Finally, we note that, at best, the Network Time Protocol synchronizes the system clock, as seen by the NTP client software, to an external reference time scale. This may not be adequate to support end-to-end traceability as we discuss below.

VI. TRACEABILITY IN FINANCIAL MARKETS

Legal time traceability in financial markets has been a concern since about 1996, when the U. S. Securities and Exchange Commission (SEC) began investigating the practices of the National Association of Securities Dealers (NASD) and the NASDAQ stock market, and found that the methods that they used to execute trades were not always in the best interests of stock market investors. As a result of a legal settlement, the NASD was required to develop an order audit trail system (OATS) so that auditors could determine whether or not trades were executed in the same order that they were received. Developing a successful OATS required synchronizing every clock involved in a stock market transaction to a common time reference. The official time reference for U. S. stock market transactions was chosen to be NIST time, and the first synchronization requirement for financial markets, OATS Rule 6953 [19], went into effect in August 1998. Since then, all major U. S. financial markets require clocks to be referenced to UTC(NIST), and to provide evidence that traceability to NIST has been established [20]. The synchronization requirement was originally just 3 s [19, 21]. It was later reduced to 1 s (a requirement that is still in effect for manual orders) [22] and then to the current 50 ms synchronization requirement that applies to all computer clocks and all automated orders. The 50 ms requirement was approved by the SEC in 2016 and first implemented in February 2017 [23, 24]. Table 1 summarizes past and current synchronization requirements for U. S. financial markets.

Table 1. A summary of U. S. financial market synchronization requirements.

Rule Number	Author	Year of Origin	Current Status	Reference Time Source	Synchronization Requirements with Respect to Reference Time Source	
OATS Rule 6953 [19]	NASD	1998	Superseded by 7430	NIST	All clocks	3 s
NYSE Rule 132A [21]	NYSE	2003	Superseded by 7430	NIST	All clocks	3 s
OATS Rule 7430 [22]	FINRA	2008	In effect	NIST	All clocks	1 s
Regulatory Notice 16-23 [23]	FINRA	2016	In effect	NIST	Computer clocks	50 ms
					Mechanical clocks	1 s
Consolidated Audit Trail National Market Plan (CAT NMS) [24]	SEC	2016	In effect	NIST	Automated orders	50 ms
					Manual orders	1 s
					Time stamp resolution	1 ms

The current European standard is the MiFID II (Market in Financial Instruments Directive). The European Securities and Markets Authority (ESMA), who is empowered by MiFID to draft regulatory technical standards, has a role similar to that of the SEC in the United States. Except for manual orders, where the 1 s requirement is identical to that in the U. S., the proposed European synchronization requirements are far more stringent. Automated orders require 1 ms synchronization and high frequency trading (HFT) orders require 0.1 ms (100 µs) synchronization, where synchronization is defined as the maximum divergence from UTC. The required time stamp resolution, which does not exceed 1 ms in the U. S., is 1 µs in Europe for HFT orders. Table 2 summarizes the synchronization requirements for European financial markets that went into effect on January 3, 2018 [25].

Table 2. A summary of European financial market synchronization requirements.

Rule Number	Author	Year of Origin	Current Status	Reference Time Source	Synchronization Requirements with Respect to Reference Time Source	
MiFID II	ESMA	2015	In effect as of January 3, 2018	UTC	Manual orders	1 s
					Automated orders, non-HFT	1 ms
					HFT	0.1 ms
					HFT time stamp resolution	1 µs

Traceability to UTC is required, but any timing laboratory that contributes to UTC (the UTC(k) laboratories) can serve as the reference time source. According to the MiFID II clock synchronization guidelines published by ESMA,

> *"systems that provide direct traceability to the UTC time issued and maintained by a timing centre listed in the BIPM Annual Report on Time Activities are considered as acceptable to record reportable events. The use of the time source of the U.S. Global Positioning System (GPS) or any other global navigation satellite system such as the Russian GLONASS or European Galileo satellite system when it becomes operational is also acceptable to record reportable events provided that any offset from UTC is accounted for and removed from the timestamp. GPS time is different to UTC. However, the GPS time message also includes an offset from UTC (the leap seconds) and this offset should be combined with the GPS timestamp to provide a timestamp compliant with the maximum divergence requirements …."*. [26]

These guidelines indicate that UTC(NIST), UTC(USNO), or UTC(USNO via GPS) can all satisfy MiFID II requirements when the appropriate level of documentation is provided and serve as the reference time source for financial transactions in Europe. The guidelines do not distinguish between a real-time use of UTC(USNO via GPS), which is a prediction of UTC(USNO) and at best require an addition to the uncertainty budget, and after-the-fact use, which could be a measurement.

The stock market requirements for synchronization apply to all of the computer clocks that are involved in time stamping financial transactions, and it should be noted that it is typically easy to synchronize a GPS clock to within 1 µs, but keeping a group of computer clocks synchronized to within 100 µs is a more challenging problem. Computer clocks are normally

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

referenced to inexpensive quartz oscillators with large instabilities and drift rates, and thus frequent synchronization to either a NTP or a Precision Time Protocol (PTP) server is required to ensure that the synchronization requirements are continuously met. The 50 ms requirement specified in the U. S. standards can still be met through periodic synchronization to NIST or USNO NTP servers via the public Internet (Section V), but the more stringent 100 µs requirement for HFT in Europe will require the continuous synchronization of NTP or PTP servers to a UTC source, and that these ideally will be located very close to the stock exchange, so that the round trip network delay between the time servers and the computers involved in stock transactions can be made as small as possible.

NIST is currently providing synchronization and establishing traceability for financial markets in the United States, Europe, and Asia with common-view GPS clocks that are disciplined to UTC(NIST), as described in Section IV. The NISTDC is located in the same data center as the stock exchange, and is used to synchronize NTP and PTP time servers that are also located in the data center. An additional service offered by NIST can now measure and verify the packets transmitted by the time servers by comparing them to the NISTDC [20].

All of the methods we have described can support the traceability of the system clock that is used to provide the time stamps for commercial or financial transactions to a UTC(*k*) time scale. However, completing the traceability chain requires assigning a measurement uncertainty to the time stamps received by the end user, which in turn requires knowledge of the latency in the link between the application that time stamps a transaction (often called the "matching engine") and the system time, which is being controlled by a separate process with its own latencies. These latencies are probably small enough to support traceability requirements at the millisecond-level, but could be too large to support microsecond-level uncertainties. Users are responsible for including the uncertainties of this final link in the traceability chain in their uncertainty analysis. These uncertainties can be especially large for applications that run in the "cloud" where there is generally a much weaker connection between the physical system clock and the time seen by an application.

VII. TRACEABILITY IN THE ELECTRIC POWER INDUSTRY

As electric power usage has increased, the electric power industry has become more and more dependent upon accurate time. Synchronized phasors, or synchrophasors, are referenced to an absolute point in time by using a common UTC time reference. The devices that perform the synchrophasor measurements are known as phasor measurement units (PMUs). A PMU measures positive sequence voltages and currents at power system substations, and time stamps each measurement. The measurement results are then sent through a network to a central site, where the time stamps are aligned, the measurements are processed, and real-time decisions are made about how to allocate power within the grid to prevent outages.

The *IEEE C37.118.1* document is the standard for synchrophasor measurements. Section 4.3 of the standard specifically mentions traceability to UTC in three places, beginning with the clause stating that "the PMU shall be capable of receiving time from a reliable and accurate source, such as the Global Positioning System (GPS), that can provide time traceable to UTC ….". Time synchronization of 26 µs corresponds to a phase angle error of 0.57 º at the 60 Hz AC line frequency, which in turn corresponds to a 1% total vector error (TVE), as defined in the standard. However, the standard indicates that:

> *"A time source that reliably provides time, frequency, and frequency stability at least 10 times better than these values corresponding to 1% TVE is highly recommended. The time source shall also provide an indication of traceability to UTC and leap second changes.*
>
> *For each measurement, the PMU shall assign a time tag that includes the time and time quality at the time of measurement. The time tag shall accurately resolve time of measurement to at least 1 µs within a specified 100 year period. The time status shall include time quality that clearly indicates traceability to UTC, time accuracy, and leap second status."* [27]

The language calling for a source providing time "at least 10 times better" than 1% TVE indicates that at least 2.6 µs synchronization is recommended. The desired accuracy is usually given as 1 µs, which corresponds to a phase angle error of only 0.022° at 60 Hz, and the time tags also require 1 µs resolution. These requirements are difficult to meet at geographically dispersed locations without GPS clocks. Therefore, in practice the time source is nearly always a GPS clock, or a PTP system that is synchronized to a GPS clock and accessed by multiple PMUs over a LAN.

VIII. TRACEABILITY OVER LEAP SECONDS AND OTHER IRREGULARITIES

Leap seconds are integer seconds that are added to UTC to maintain the difference between UTC and UT1 less than ±0.9 s. They are always added as the last second of the last day of a month, with the last days of June and December preferred. The leap seconds can also be added at other times if needed. A leap second is inserted after 23:59:59 UTC, and its official name is 23:59:60. Since leap seconds are inserted in the UTC time scale, they occur late in the afternoon in the US Pacific time zone and near noon in Asia and Australia.

Digital clocks in general, and computer clocks in particular, cannot display a time corresponding to 23:59:60, and various ad-hoc techniques must be used to define a time stamp during a leap second. Many systems, including the NIST network time servers, effectively stop the clock during a leap second, and transmit a time corresponding to 23:59:59 a second time. A similar technique repeats the time stamp corresponding to 00:00:00 of the next day a second time. (This technique has the correct long-term behavior but puts the extra second in the wrong day.) Both of these techniques have an ambiguity in the time stamp transmitted in the NTP format, since the format cannot distinguish between the first and second time stamps with the same integer value. This problem introduces an ambiguity in the time-ordering of events in the vicinity of the leap second. (For example, the NIST time servers will receive about ~900 000 requests for time during the two seconds with identical time stamps of 23:59:59.) Some operating systems "solve" the leap second problem by ignoring them altogether, and all systems are susceptible to software bugs.

Leap seconds introduce a numerical discontinuity into a time interval that includes the leap second, so that real-time systems, such as GPS system time, do not use them. GPS (and other GNSS, with the exception of GLONASS) transmit the integer-second offset between system time and UTC as part of the navigation message, shown in Eq. (2). User equipment can use this parameter to translate between system time and UTC, but this does not solve the problem of the ambiguity in the name assigned to the leap second.

Some corporations, in an attempt to minimize the impact on their systems and eliminate the discontinuity, have implemented "smears", that slow down their clocks for a period around the time of the leap second insertion [28]. This method has the advantage that the time stamps are monotonically increasing even in the vicinity of the leap second, but it has an error of order ±0.5 s with respect to the definition of UTC during the interval of the frequency adjustment. In addition, there is no standard method for applying this frequency adjustment, so that different implementations may disagree among themselves in addition to the time error with respect to UTC. Metrological traceability can be maintained through a leap second only if the user is able to program the systems to keep track of the extra second, or the smear, and take it into account.

IX. TRACEABILITY AND LATENCY

The official definitions of traceability do not address the issue of the non-zero latency between the measurement and the determination of the uncertainty. This often-small latency can convert even a direct measurement into a prediction awaiting confirmation. In many cases the assumption that the difference between the actual value and the prediction of a UTC(k) is of zero mean and has a near-Gaussian distribution can be justified on the basis of past data. The assumption can also be verified after-the-fact, and any user whose traceability data are being carefully scrutinized would be well-advised to take the verification into consideration. For example, a user of GPS via direct access should confirm that the results shown in Figure 2 continue as shown.

X. SUMMARY

Time signals, including those transmitted by GPS satellites and network time servers, can be used to establish legal metrological traceability to UTC through a UTC(k) time scale. These signals have small enough uncertainties to meet industrial synchronization and traceability requirements, but users are responsible for having sufficient evidence to prove that their requirements are being met. A critical part of this evidence is the ability to demonstrate that an unbroken chain of calibrations back to UTC through a UTC(k) laboratory exists, and that each link of the traceability chain has a documented measurement uncertainty.

Matsakis, Demetrios; Levine, Judah; Lombardi, Michael.
"Metrological and legal traceability of time signals."
Paper presented at 2018 Precise Time and Time Interval (PTTI) Meeting, Reston, VA, United States. January 29, 2018 - February 1, 2018.

IX. REFERENCES

[1] Joint Committee for Guides in Metrology (JCGM), 2008, "International vocabulary of metrology – Basic and general concepts and associated terms (VIM)," *JCGM 200*, 3rd edition, p. 29.

[2] International Telecommunication Union (ITU), 2013, "Glossary and definitions of time and frequency terms," *ITU-R TF.686-3*, p. 16.

[3] Resolution 1 of the *13th Conference Generale des Poids et Mesures (CGPM)*, 1967.

[4] G. Panfilo, 2016, "The coordinated universal time," *IEEE Instrum. Meas. Mag.* 19, 28–33.

[5] J. Levine, 1999, "Introduction to time and frequency metrology," *Rev. Sci. Instrum.*, 70, 2567-2596.

[6] The U.S. Securities and Exchange Commission provides information about Rule 613 here: https://www.sec.gov/divisions/marketreg/rule613-info.htm

[7] Joint Committee for Guides in Metrology (JCGM), 2008, "Evaluation of measurement data – Guide to the expression of uncertainty in measurement," *JCGM 100*.

[8] U. S. Code, Title 15, Chapter 6, Subchapter 9, section 261, https://www.law.cornell.edu/uscode/text/15/261

[9] DoD Instruction 4650.06: http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/465006p.pdf. DoD users may also consult the Chairman of the Joint Chiefs of Staff's Instruction CJCSI 6130.01F.

[10] Global Positioning Systems Directorate, 2013, "Navstar GPS Space Segment/Navigation User Interfaces," *Interface Specification IS-GPS-200H*.

[11] See also https://www.gps.gov/technical/icwg/.

[12] M. Lombardi, 2016, "Evaluating the Frequency and Time Uncertainty of GPS Disciplined Oscillators and Clocks," *NCSLI Measure: The Journal of Measurement Science*, 11, 30-44.

[13] M. Lombardi, 2010, "A NIST Disciplined Oscillator: Delivering UTC(NIST) to the Calibration Laboratory," *NCSLI Measure: The Journal of Measurement Science*, 5(4), 46-54.

[14] J. Sherman and J. Levine, 2016, "Usage Analysis of the NIST Internet Time Service," *Journal of Research of the National Institute of Standards and Technology*, 121, 33-46.

[15] J. Levine, 2016, "An Algorithm for Synchronizing a Clock When the Data are Received Over a Network With an Unstable Delay," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63, 561-570.

[16] D. Matsakis, 2013, "Network Time Protocol (NTP) Accuracy as Seen by the Users," in Proceedings of the 45th Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 2-5 December 2013, Bellevue, Washington, USA, Virginia, USA, 173-177.

[17] J. Levine, 2017, "The UTI and UTC Time Services Provided by the National Institute of Standards and Technology," Chapter 17 in *The Science of Time 2016*, Astrophysics and Science Proceedings, 50, 127-139.

[18] A. Novick and M. Lombardi, 2015, "Practical Limitations of NTP Time Transfer," *Proceedings of the 2015 Joint Conference of the IEEE International Frequency Control Symposium and the European Frequency and Time Forum*, 12-16 April 2015, Denver, Colorado, 570-574.

[19] National Association of Securities Dealers (NASD), OATS Reporting Technical Specifications (1998).

[20] M. Lombardi, A. Novick, B. Cooke, and G. Neville-Neil, 2016, "Accurate, Traceable, and Verifiable Time Synchronization for World Financial Markets," *Journal of Research of the National Institute of Standards and Technology*, 121, 436-463.

[21] United States Securities and Exchange Commission (SEC), "Self-Regulatory Organizations; Order Approving Proposed Rule Change by the New York Stock Exchange, Inc. Relating to Order Tracking," Release No. 34-47689, File No. SR-NYSE-99-51 (2003).

[22] Financial Industry Regulatory Authority (FINRA), OATS Reporting Technical Specifications (January 11, 2016).

[23] Financial Industry Regulatory Authority (FINRA), "Clock Synchronization," *FINRA Regulatory Notice 16-23* (July 2016).

[24] United States Securities and Exchange Commission (SEC), "Joint Industry Plan; Order Approving the National Market System Plan Governing the Consolidated Audit," Release No. 34-79318, File No. 4-698, (November 15, 2016).

[25] "Commission Delegated Regulation (EU) 2017/574 of 7 June 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council with regard to regulatory technical standards for the level of accuracy of business clocks," 2017, *Official Journal of the European Union*, March 2017, L 87/148-151.

[26] European Securities and Markets Authority (ESMA), "Guidelines: Transaction reporting, order record keeping and clock synchronisation under MiFID II," *ESMA/2016/1452* (corrected on July 8, 2017).

[27] IEEE Power & Energy Society, "IEEE Standard for Synchrophasor Measurements for Power Systems," *IEEE C37.118.1-2011*, 2011.

[28] C. Pascoe, 2011, "Time, technology, and leaping seconds," *Google Official Blog*, https://googleblog.blogspot.com/2011/09/time-technology-and-leaping-seconds.html

MRS Fall 2017

Observation of triplet level crossing in single crystal organic transistors using magneto conductance at room temperature

H.- J. Jang, E. G. Bittle, Q. Zhang, C. A. Richter, D. J. Gundlach

Organic semiconductors provide a unique set of properties that provide for the manufacture of large and flexible LED screens and photovoltaic arrays. In order to lower the operating voltages of organic LEDs (OLEDs) and improve efficiency above the Shockley-Queisser limit in organic photovoltaic (OPV) devices, quantum processes such as singlet fission/triplet fusion can be exploited. Singlet fission, where two triplet excitons (or polaron pairs) are generated from one photon, may help to boost the solar conversion efficiency of these van der Waals bounded materials. In the reverse process, triplets created by charge carriers injected through electrical contacts in OLEDs are suggested to be upconverted to higher energy singlets and reduce the on-voltage. Though singlet fission/triplet fusion processes show promise in enhancing efficiency, understanding of triplet and singlet state control in electronic devices where paired charges can recombine at defect and interface sites must be further explored.

In order to better understand the tie between singlet fission/triplet fusion and its related multiexcitonic intermediate state process, we present results from a study of magnetoconductance of single crystal tetracene field effect transistors. We find that we can tune the amount of current in the transistor by changing the magnitude of an applied magnetic field under light illumination. In addition, results show that the transistor magnetocurrent dips in performance at around 10mT and 42 mT at a certain applied field angle, which can be understood by considering sub-energy level crossings of paired triplet states due to the Zeeman effect and magnetic dipolar interaction. These dips in performance illustrate the direct correspondence between quantum processes and device performance, and demonstrate the possibility of OLED and OPV device control by using magnetic fields.

Bittle, Emily; Jang, Hyuk-Jae; Zhang, Qin; Richter, Curt; Gundlach, David.
"Observation of triplet level crossing in single crystal organic transistors using magneto conductance at room temperature."
Paper presented at 2017 Materials Research Society Fall Meeting, Boston, MA, United States. November 26, 2017 - December 1, 2017.

Josephson Arbitrary Waveform Synthesizer as a Reference Standard for the Measurement of the Phase of Harmonics in Distorted Signals

D. Georgakopoulos*, I. Budovsky*, S. P. Benz† and G. Gubler††

*National Measurement Institute Australia, 36 Bradfield Road, Lindfield NSW 2070 Australia
dimitrios.georgakopoulos@measurement.gov.au

† National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA

†† D I Mendeleyev Institute for Metrology, Moskovsky pr. 19, St Petersburg, 190005, Russia

***Abstract* — We use the Josephson arbitrary waveform synthesizer to provide traceability for the phase of the harmonics, relative to its fundamental, of a distorted waveform. For distorted waveforms with rms values in the range from 0.154 V to 0.2 V and harmonic magnitudes from 5% to 40 % of the fundamental, our system can generate odd harmonics up to the 39th with phase uncertainties from 0.002° to 0.010° (*k*=2.0), depending on the harmonic number and harmonic magnitude.**

***Index Terms* — Josephson junction array, measurement standards, measurement techniques, phase measurement, quantum voltage standards, spectrum analysis.**

I. Introduction

Harmonic analysis is used in electricity networks, communications, characterization of systems and materials, acoustics and vibration. While the traceability of harmonic magnitude measurements to ac-dc transfer measurement standards is well established (see for example Ref. [1]), there is a gap in the traceability for the phase of harmonics relative to the fundamental.

Budovsky [2] proposed the use of a Josephson arbitrary waveform synthesizer (JAWS) for the calibration of the phase of harmonics of a distorted signal relative to the fundamental. Based on [2], we use the JAWS to generate precisely distorted waveforms that contain harmonics of known magnitude and phase. The primary application of this work is to provide traceability for the phase of the harmonics for power analyzers used in power systems. Target uncertainties for this application range from 0.002° to 0.010°, depending on the relative harmonic magnitude and the harmonic number.

II. System Description

The JAWS system (Fig. 1) contains a continuous wave generator, a tertiary pattern generator, two broadband radio frequency (rf) amplifiers, two arbitrary waveform generators (AWGs), two voltage-to-current converters and two NIST Josephson junction arrays (JJAs) [3]. The inputs and outputs of the rf amplifiers are connected through a combination of dc-blocks and attenuators (shown as capacitors in Fig. 1), optimized for the maximum step current margins of the JJAs.

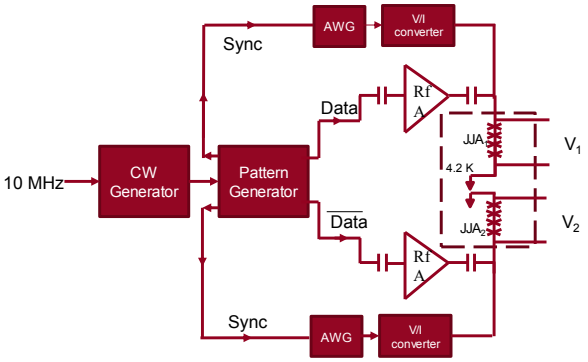


Fig. 1. Block diagram of the JAWS system

A 24-bit flexible resolution digitizer is used to check the quantization of the JJAs.

The JJAs are programmed to produce two arbitrary waveforms of the same harmonic content and rms values with a relative phase difference of 180°. The memory size of the pattern generator limits the lowest fundamental frequency to approximately 56 Hz.

III. Experimental Results

In view of the errors that can occur in a practical realization of a JAWS [4,5], to evaluate the performance of the JAWS as a harmonic phase standard it is important to show that (a) the phase of the generated signals is independent of the specific JJA, (b) that the biasing electronics do not affect the operation of the system, and (c) that the phase of each generated harmonic does not depend on the total harmonic content. In this section we present several experiments aimed at proving that these conditions are met. In these experiments we used two digital sampling systems as transfer standards – one based on two precision digital sampling multimeters [6] (Digitizer1) and the other based on the National Instruments PXI5922 (Digitizer2).

First, we generated two single-frequency sine waves of 0.158 V rms with a nominal phase difference of 180° and used Digitizer1, whose internal phase shift was nulled prior to the

Georgakopoulos, Dimitrios; Budovsky, Ilya; Benz, Samuel; Gubler, G..
"Josephson Arbitrary Waveform Synthesizer as a Reference Standard for the Measurement of the Phase of Harmonics in Distorted Signals."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

measurement, to measure the phase difference between these two sinewaves. The disagreement was 0.00003° at 60 Hz and less than 0.0003° up to 1 kHz.

Next the JAWS was used to generate a number of precisely distorted waveforms with a fundamental frequency of 60 Hz and an rms value of 0.158 V. These precisely distorted waveforms contained the 3rd, 5th, 7th, 9th, 11th, 23rd, 31st and 39th harmonics. These harmonics were selected as a compromise between (a) using the harmonics commonly found in power systems and referenced in power quality documentary standards and (b) keeping the signal-to-noise ratio of the harmonics relative to the noise floor of the spectrum analysis high enough to allow phase resolution better than 0.001°. Relative to the fundamental, the magnitude of each harmonic was 10 % and the phase was 90°. Digitizer1 was first used to measure waveforms containing the fundamental and one of the above harmonics, and then a waveform with the fundamental and all of the above harmonics (Fig 2). The difference between the two sets of measurements did not exceed 0.001°.

Several measurements were performed to study the influence of the biasing electronics, which includes the continuous wave generator, the pattern generator, the rf amplifiers, the voltage-to-current converters and the arbitrary waveform generators. First we calibrated Digitizer2 with two different JJAs from the same chip, driven by the same biasing electronics, using the all-harmonics waveform specified above. The results of the two measurements agree within 0.001° and the measured errors of the digitizer are within 0.001° of nominal (Fig. 3a). Then the two JJAs were used to generate the same waveform, but driven by different biasing electronics (Fig. 3b). For all the harmonics up to the 23rd, the results agree within 0.001°. As the harmonic number increased the disagreement increased as well but was still within the uncertainty of the measurement. The error bars in Fig. 3 indicate the standard deviation of the measurement.

Finally, a number of measurements were conducted with varying distortion, phase angle, helium level, and JJA output cable length. These results and the uncertainty analysis will be reported at the conference.

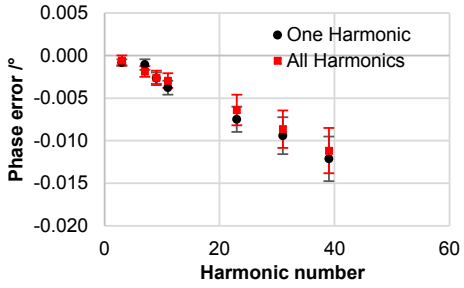


Fig. 2 Phase error of Digitizer1 measured with one harmonic at a time and all harmonics simultaneously. The error bars indicate the standard deviation of the measurement.

III. Conclusion

The results presented in this paper show that the JAWS gives consistent results as a standard for the phase of harmonics of a distorted signal for different JJAs and the biasing electronics. The uncertainty analysis, to be presented at the conference, suggests that the system can generate distorted signals with phase uncertainty of the harmonics from 0.002° to 0.010° (k=2.0), depending on the harmonic magnitude and harmonic number.

References

[1] I. Budovsky and G. Hammond, "Precision measurement of power harmonics and flicker," *IEEE Trans. Instrum. Meas.,* vol. 4, no. 4, pp. 483 – 487, April 2005.

[2] I. Budovsky, "Development of a harmonic phase measurement system traceable to a Josephson waveform synthesizer," EURAMET TCEM sub-committee DC and Quantum Metrology Annual Meeting, MIKES, Espoo, Finland, 2017.

[3] S. P. Benz, S.B. Waltman, A. E. Fox, P.D. Dresselhaus, A. Rufenacht, J. M. Underwood, L.A. Howe, R.E. Schwall, and C.J. Burroughs, "One-volt Josephson arbitrary waveform synthesizer," *IEEE Trans. Supercond.,* vol. 25, no. 1, p. 1300108, February 2015.

[4] R. P. Landim, S.P. Benz., P.D. Dresselhaus and C. J. Burroughs, "Systematic-error signals in the ac Josephson voltage standards: measurement and reductions," *IEEE Trans. Instrum. Meas.,* vol. 57, no. 6, pp. 1215 – 1220, June 2007.

[5] N.E. Flowers-Jacobs, A. Rufenacht, A.E. Fox, P.D. Dresselhaus, and S. P. Benz, "2 volt pulse-driven Josephson arbitrary waveform synthesizer," CPEM 2016 Conf. Digest, pp.1 – 2, July 2016.

[6] G. Gubler and E.Z. Shapiro, "Implementation of sampling measurement system for new VNIIM power standard," CPEM 2012 Conf. Digest, pp.294, July 2012.

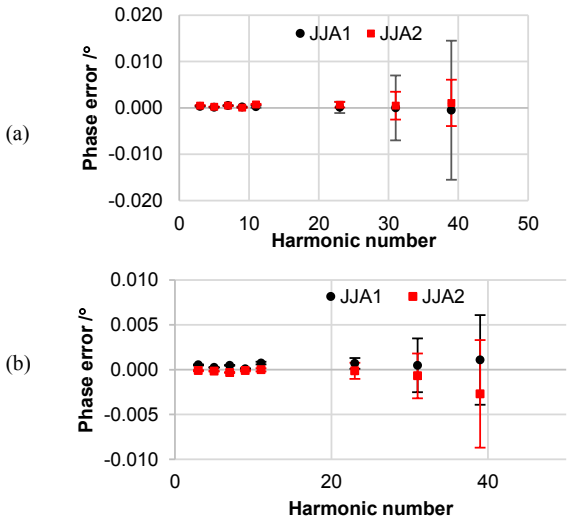


Fig. 3 Phase error of Digitizer2 measured using two different JJAs driven by (a) the same biasing electronics and (b) different biasing electronics.

Georgakopoulos, Dimitrios; Budovsky, Ilya; Benz, Samuel; Gubler, G..
"Josephson Arbitrary Waveform Synthesizer as a Reference Standard for the Measurement of the Phase of Harmonics in Distorted Signals."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Characterization of a Dual Josephson Impedance Bridge

Frédéric Overney[1], Nathan E. Flowers-Jacobs[2], Blaise Jeanneret[1], Alain Rüfenacht[2],
Anna E. Fox[2], Paul D. Dresselhaus[2] and Samuel P. Benz[2]
[1]Federal Institute of Metrology METAS, Lindenweg 50, 3003 Bern-Wabern, Switzerland
[2]National Institute of Standards and Technology, Boulder, CO 80305, USA

***Abstract*—This paper describes a dual Josephson impedance bridge capable of comparing any two impedances, that is, with any amplitude ratio and relative phase, over a wide range of frequency. A new, more compact, design has been achieved by mounting the two Josephson Arbitrary Waveform Synthesizers (dual JAWS) side-by-side in a single cryoprobe. We measured the crosstalk between the two JAWS sources and show that, by interchanging the impedances within the bridge, the effect of the crosstalk between JAWS sources can be reduced to a negligible level.**

***Index Terms*—Impedance comparison, AC Josephson voltage standard, Josephson Arbitrary Waveform Synthesizer, AC coaxial bridge.**

I. Introduction

The comparison of impedance standards using transformer-based bridges is widely used in national metrology institutes (NMIs) and allows the calibration of impedances with the highest accuracy in the audio frequency range [1], [2]. The major drawback of such bridges is that impedances can be compared only at a set of specific predefined magnitude ratios (typically 1:1 or 1:10) and relative phases (typically 0° or ±90°).

This limitation has been overcome by replacing the transformer with two ac Josephson voltage standards (also known as dual Josephson Arbitrary Waveform Synthesizers or dual JAWS sources), which generate accurate and stable voltages with a completely programmable magnitude ratio and relative phase. The first realization of an impedance bridge based on two JAWS circuits demonstrated the capability to compare any two kinds of impedances over the whole audio frequency range [3], [4].

Since the first realization of this Dual Josephson Impedance Bridge (DJIB) in 2016 [3], the system has been further optimized: the electronics [5] have been modified to make the system more compact [6] and the JAWS design has been improved to fit both sources in the same cryoprobe, which is cooled in a single helium Dewar. The preliminary description of this new dual source and the characterization of the new DJIB are presented in this summary.

II. Dual JAWS setup

Fig. 1 represents schematically the wiring of the two JAWS sources located side-by-side in the cryoprobe. For clarity, the microwave circuitry is omitted. Each JAWS is composed of four arrays of 12 810 Josephson junctions (JJs) with critical currents of 10 mA. The JJs in each JAWS are driven by a single pulse generator channel; a four-way split is accomplished on-chip using two layers of Wilkinson dividers [6].

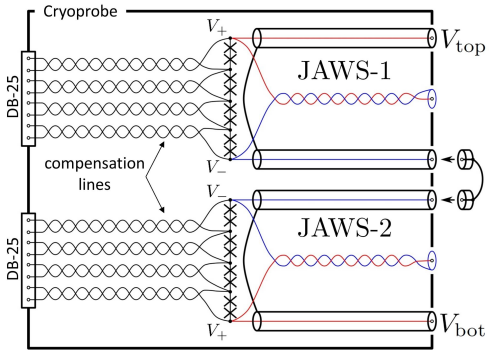


Fig. 1. Schematic representation of the low-frequency wiring of the two JAWS sources inside the cryoprobe. The grounding scheme of the circuit can be changed by connecting the grounding plugs shown near the V_- outputs on the right side of the figure.

For each JAWS, the inner conductor of each of two coaxial cables brings the audio-frequency quantum-accurate Josephson voltage (V_+ and V_-) from the chip to the top of the cryoprobe. The outer conductors of these coaxial cables are isolated from the cryoprobe and connected together near the chip. Each JAWS also has a pair of twisted wires connected in parallel with the coaxial cables for system testing (colored wires in Fig. 1). Four supplementary twisted-pair wires are connected from DB-25 connectors at the top of the cryoprobe to the four JJ arrays of each JAWS to provide the compensation current needed to generate voltages > 200 mV.

When the dual JAWS sources are implemented in the DJIB, the V_- outputs of the two JAWS circuits are connected together and shorted to ground using the grounding plugs shawn in Fig. 1. The V_+ outputs supply the quantum-accurate voltages $V_{\rm top}$ and $V_{\rm bot}$ to the bridge.

III. Crosstalk

As described above, the two JAWS sources are located side by side. Although this configuration allows a more compact design, the question of crosstalk between the two JAWS circuits must be investigated in detail.

A. Effect of crosstalk on the DJIB

Fig. 2 shows schematically the effect of JAWS crosstalk on the DJIB. Once the bridge is balanced, i.e., no current is

Flowers-Jacobs, Nathan; Jeanneret, Blaise; Overney, Frederic; Rufenacht, Alain; Fox, Anna; Dresselhaus, Paul; Benz, Samuel.
"Characterization of a Dual Josephson Impedance Bridge."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

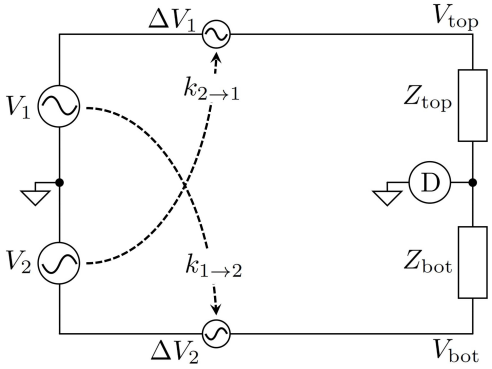


Fig. 2. Schematic representation of the cross-talk in the DJIB.

flowing through the detector D, the impedance ratio $Z_{\rm bot}/Z_{\rm top}$ is determined by the voltage ratio $V_{\rm bot}/V_{\rm top}$. The crosstalk will induce a deviation in $V_{\rm bot}$ from the JAWS signal V_2 by a quantity $\Delta V_2 = k_{1\to2}V_1$ where $k_{1\to2}$ is the crosstalk coefficient. Similarly, $V_{\rm top}$ will differ from V_1 by $\Delta V_1 = k_{2\to1}V_2$.

We first consider the case of symmetric crosstalk $k_{1\to2} = k_{2\to1} = k$, where k is complex and $|k| << 1$. The balance equation is then given by:

$$\frac{Z_{\rm bot}}{Z_{\rm top}} = -\frac{V_{\rm bot}}{V_{\rm top}} = -r\frac{1+k\frac{1}{r}}{1+kr} \approx -r\Big(1+k(\frac{1}{r}-r)\Big), \quad (1)$$

where $r = V_2/V_1$ is the ratio of quantum-accurate reference voltages. Eq. 1 clearly shows that the crosstalk has a negligible effect when comparing impedances of the same kind (R to R or C to C) in a 1:1 ratio (i.e.: $r \approx 1$). However, when the ratio differs from 1, the effect of the crosstalk must be taken into account.

It has been shown [3] that one of the unique and important properties of the DJIB is the ability to repeat the balance of the bridge with the two impedance standards inverted, even with $r \neq 1$. The inverted balance of the bridge leads to the following equation:

$$\frac{Z_{\rm top}}{Z_{\rm bot}} = -\tilde{r}\frac{1+k\frac{1}{\tilde{r}}}{1+k\tilde{r}}, \quad (2)$$

where $\tilde{r}$ is the new generated voltage ratio $\tilde{V_2}/\tilde{V_1}$. Combining Eq. 1 and Eq. 2, the impedance ratio is finally given by

$$\frac{Z_{\rm top}}{Z_{\rm bot}} \approx \sqrt{\frac{r}{\tilde{r}}\Big(1+k^2(\frac{1}{r^2}-r^2)\Big)}, \quad (3)$$

assuming that $\tilde{r} \approx 1/r$. Eq. 3 clearly shows the advantage of performing both the direct and reverse measurements: the impact of the crosstalk is reduced to second order.

B. Measurement of the crosstalk coefficients

To determine the crosstalk coefficient k, the voltage ΔV_2 was measured at frequencies between 1 kHz to 50 kHz, with the amplitude of V_2 set to zero and the rms amplitude of V_1 set to 1 V. The amplitude $|\Delta V_2|$ increases linearly with frequency and reaches an rms amplitude between 1.4 μV and 130 μV at 50 kHz, depending on the DJIB grounding scheme. The smallest crosstalk coefficient of 1.4 μV was obtained when both JAWS sources were completely floating, i.e., when the two short-circuit plugs of Fig. 1 were not connected together.

The origin of the crosstalk appears to be the capacitive current that flows from one JAWS source to the other, either between on-chip elements or between various twisted-pair wires connecting the JAWS chip to the top of the cryoprobe. This capacitive current, $j\omega CV_1$, needs to return to ground through the impedance of the link between JAWS-2 and JAWS-1, $z = R_s + j\omega L_s$. It produces a voltage drop $\Delta V_2 = j\omega CV_1 z$. The impedance z includes the on-chip impedance of the JJ arrays and superconducting wiring plus the series impedance of the coaxial cables. Because the same capacitance C and impedance z are involved in the reverse coupling, the hypothesis of a symmetric crosstalk coefficient is fully justified. We used a four-wire measurement to determine the total output impedance, $R_s = 2.5\ \Omega$ and $L_s = 2.6\ \mu$H, which implies that a coupling capacitance between the JAWS sources of $C = 160$ pF would be required to explain the maximum measured crosstalk of $\Delta V_2 = 130\ \mu$V at 50 kHz.

This explanation for the crosstalk is further supported by the fact that the phase of the measured voltage ΔV_2 is shifted by roughly $90°$ relative to the phase of V_1. Moreover, when the two JAWS circuits are completely floating, the capacitive current vanishes because there is no path back to the source.

IV. Conclusion

Mounting the two JAWS sources side-by-side in a single cryoprobe results in a simple, compact, dual Josephson source. However, this configuration requires careful measurement of crosstalk and accounting for the effect of the crosstalk on the calculated voltage ratio, especially when the ratio differs from 1. By combining both direct and reverse measurements, the effect of the crosstalk on the calculated voltage ratio is reduced to second order and does not significantly contribute to the ratio, even though the coupling between the two JAWS circuits is as large as 160 pF. The next step for the DJIB is to detremine a full uncertainty budget and directly compare it to a high-accuracy transformer-based bridge.

References

[1] B. Hague and T. R. Foord, *Alternating Current Bridge Methods*. Pitman Publishing, 1971.

[2] S. A. Awan, B. Kibble, and J. Schurr, *Coaxial Electrical Circuits for Interference-Free Measurements*, ser. IET Electrical Measurement Series. Institution of Engineering and Technology, Jan 2011.

[3] F. Overney *et al.*, "Josephson-based full digital bridge for high-accuracy impedance comparisons," *Metrologia*, vol. 53, no. 4, pp. 1045–1053, Aug 2016.

[4] S. Bauer *et al.*, "A novel two-terminal-pair pulse-driven Josephson impedance bridge linking a 10 nF capacitance standard to the quantized Hall resistance," *Metrologia*, vol. 54, no. 2, pp. 152–160, Apr 2017.

[5] S. P. Benz and S. B. Waltman, "Pulse-Bias Electronics and Techniques for a Josephson Arbitrary Waveform Synthesizer," *IEEE Transactions on Applied Superconductivity*, vol. 24, no. 6, pp. 1–7, Dec 2014.

[6] N. E. Flowers-Jacobs, S. B. Waltman, A. E. Fox, P. D. Dresselhaus, and S. P. Benz, "Josephson Arbitrary Waveform Synthesizer With Two Layers of Wilkinson Dividers and an FIR Filter," *IEEE Transactions on Applied Superconductivity*, vol. 26, no. 6, pp. 1–1, sep 2016.

Flowers-Jacobs, Nathan; Jeanneret, Blaise; Overney, Frederic; Rufenacht, Alain; Fox, Anna; Dresselhaus, Paul; Benz, Samuel.
"Characterization of a Dual Josephson Impedance Bridge."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Radio-Frequency Waveform Synthesis with the Josephson Arbitrary Waveform Synthesizer

Justus A. Brevik*, Christine A. Donnelly*†, Nathan E. Flowers-Jacobs*, Anna E. Fox*, Peter F. Hopkins*, Paul D. Dresselhaus*, and Samuel P. Benz*

*National Institute of Standards and Technology, Boulder, CO, 80305, USA
justus.brevik@nist.gov

†Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

***Abstract* — We have measured the output power versus synthesis frequency (transfer function), up to 100 MHz, of a modified Josephson arbitrary waveform synthesizer. A 50 Ω resistor was integrated within the Josephson junction array circuit to match the nominally zero array impedance to a 50 Ω transmission line and load. The resistor significantly reduces ripple in the transfer function due to reflections from impedance mismatches, but it reduces the output voltage by a factor of two. The measured power is constant within 0.01 dB up to 10 MHz, above which it deviates by no more than 0.15 dB.**

***Index Terms* — Digital-analog conversion, Josephson junction arrays, signal synthesis, standards, superconducting device measurements, superconducting integrated circuits, voltage measurements.**

I. Introduction

The Josephson arbitrary waveform synthesizer (JAWS) has been used to generate quantum-accurate waveforms with fundamental frequencies up to 1 MHz [1]. We have recently begun developing a JAWS system that can synthesize waveforms at frequencies ranging from the kilohertz to the gigahertz range and drive 50 Ω loads. Ideally, this system would be able to do so while providing a quantum-accurate output voltage at room temperature over the frequency range of interest. As a first step toward gigahertz-frequency waveforms, we have synthesized single-tone sinusoids with frequencies up to 100 MHz.

Although it is possible to measure a quantum-accurate voltage across the Josephson junction (JJ) array – with each junction producing a quantized voltage pulse for every bias pulse – the accuracy of the output voltage is degraded by the circuit that connects the array to the room temperature load, and by the load itself. The main sources of voltage error are: (1) impedance mismatches between the superconducting circuit, cabling and the load, which lead to standing-wave voltage deviations [2]; (2) loss in the cables that connect the superconducting circuit to room temperature circuits; and (3) a frequency-dependent inductive voltage error [1].

In order to drive 50 Ω loads, two modifications must be made to the standard audio-frequency JAWS circuit so that it is properly matched to those loads. In the audio-frequency JAWS circuit, a pair of filtered taps that connect to twisted-pair leads are used to measure the voltage generated across the array. The taps separate the desired low-frequency signal from the high-frequency pulse bias. The filters and twisted-pair leads have insufficient bandwidth for waveform synthesis up to 100 MHz, and do not have 50 Ω impedance. We have replaced the filtered, twisted-pair lines with a 50 Ω coaxial transmission line. For this first design, the filters have been omitted, so there is no on-chip separation of the output signal and pulse bias.

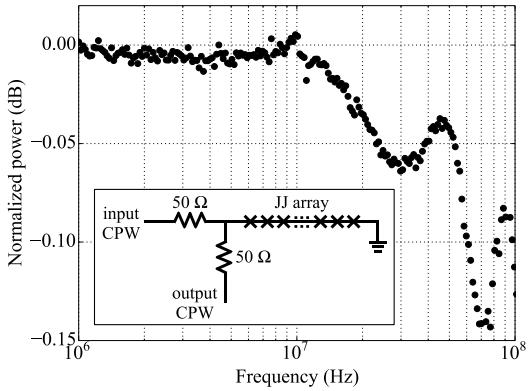


Fig. 1. Measured output power versus synthesis frequency, normalized to the power at 1 MHz. *Inset:* The modified superconducting circuit used in this version of the JAWS system.

Josephson junction arrays have an output impedance of roughly 0 Ω. This means the audio-frequency JAWS circuit is not properly impedance-matched to any finite load. When low-frequency waveforms ($\lesssim$ 100 kHz) are synthesized, this impedance mismatch has a negligible effect on the voltage accuracy. However, for synthesis frequencies above a few megahertz, the standing waves created by reflections between the array and load lead to ripple in the output power versus synthesis frequency (transfer function). One solution, which we explore in this work, is to add a 50 Ω impedance-matching resistor to the JJ array circuit. Although this method significantly improves the accuracy at high-frequency by reducing the transfer function ripple, the voltage drop across the resistor reduces the output voltage by a factor of approximately two. Even with this improvement, there may be residual standing waves due to impedance mismatches between the matching resistor, cabling to room temperature, and the load, as well as parasitic impedances in the load and throughout the measurement circuit. There will also be loss of

Brevik, Justus; Donnelly, Christine; Flowers-Jacobs, Nathan; Fox, Anna; Hopkins, Peter; Dresselhaus, Paul; Benz, Samuel.
"Radio-frequency Waveform Synthesis with the Josephson Arbitrary Waveform Synthesizer."
Paper presented at 2019 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

accuracy in the output voltage with frequency due to the frequency-dependence of the load and the loss in the output cables.

Another source of error that must be addressed is the frequency-dependent inductive error signal. The high-frequency current bias in a zero-compensation pulse bias has some remaining signal at the fundamental synthesis frequency that drives the distributed inductance of the coplanar waveguide (CPW) in the JJ array [1]. This creates a voltage error signal at the synthesis frequency, with a magnitude that is proportional to that frequency. For high-frequency waveform synthesis this error, which adds in quadrature to the desired output voltage, is expected to dominate.

II. Output Power Versus Frequency Measurement

We measured the output power versus synthesis frequency of the simple JJ array circuit shown in the inset of Fig. 1. In this circuit, a CPW routes the high-frequency bias signal to an array of 102 JJs embedded in the CPW, in 34 stacks with three JJs each. The end of the JJ array is terminated to the ground plane. A 50 Ω resistor is embedded in the input CPW leading to the array to absorb unwanted reflections of the pulse bias from the short at the end of the array. Another 50 Ω resistor is embedded in the output CPW tap, thereby placing an impedance-matched resistor in series with the JJ array.

The pulse-bias input and voltage output CPWs were ribbon-bonded to a controlled-impedance circuit board, which routed those channels through two 1.3 m, 40 GHz coaxial cables to room temperature. All of the above hardware was housed in a cryoprobe, which was submerged in liquid helium for the measurements. From the top of the cryoprobe, the input channel was connected to a series of inner-only dc blocks, two 3 dB attenuators, and a bias tee. The pulse-bias signal was provided by a commercial 65 gigasamples-per-second, 8-bit arbitrary waveform generator (RF-AWG) and amplified by a commercial 50 GHz modulator driver. The cryoprobe output channel was connected to a spectrum analyzer (SA) through a 20 GHz coaxial cable, with the shortest possible length of 10 cm. The SA had an input impedance of 50 Ω.

The frequency-dependence of the system output power was measured by synthesizing single-tone waveforms ranging in frequency from 1 MHz to 100 MHz. For each measurement, the sinusoidal waveform was converted into a three-level delta-sigma modulated pulse bias, which in turn was converted to a five-level zero-compensation code using the method described in [1]. The amplitude of the compensation pulses in the five-level code was then optimized by measuring the feedthrough signal versus amplitude at several frequencies across the range, as explained in [1]. A quantum locking range of 3.5 mA was first obtained for a 1 kHz waveform. The locking ranges were then verified over the entire synthesis frequency range, to ensure that the output voltage across the array was quantum-accurate and that any deviations were generated between the array and load. An automated program was used to program the RF-AWG with the desired single-tone pulse bias, set the SA to the proper center frequency, and acquire a trace for each frequency step with 1 kHz video and resolution bandwidths. The measurement was repeated with a 6 dB attenuator on the pulse-bias input to the array to quantify the inductive error signal versus synthesis frequency, as explained in [1].

The relative transfer function measurement, spanning 1–100 MHz, is shown in Fig. 1. The transfer function has been normalized to the power at 1 MHz, because no absolute calibration of the SA was performed and the impedance of the matching resistor had not been characterized. The frequency response of the output power is constant within 0.01 dB up to ~10 MHz, but begins to decrease and oscillate with increasing frequency. The transfer function shows a ~0.15 dB decrease in output power from 10 MHz to 100 MHz. When the 10 cm coaxial cable connecting the cryoprobe to the SA was replaced with a similar 1 m cable, the power dropped by ~0.25 dB over the range 7–100 MHz, suggesting this deviation was caused by frequency-dependent loss in the cable. The transfer function ripple, which is ~0.1 dB peak-to-peak, showed a shift in frequency periodicity with cable length. This is consistent with a standing wave between the superconducting circuit and the SA. The measurement of the inductive error signal versus synthesis frequency (not shown) increased with synthesis frequency as expected and reached a level of -37 dB relative to the fundamental tone at 100 MHz.

VI. Conclusion

A simple, impedance-matching JJ circuit has been used to measure the frequency dependence of the output power of a high-frequency replacement to the standard audio-frequency JAWS circuit. The transfer function shows flat frequency response up to 10 MHz, and a deviation of roughly 0.15 dB from 10 MHz to 100 MHz. We expect that the inductive error signal will become the dominant source of error at synthesis frequencies above 10 MHz. An upcoming circuit revision will include an on-chip superconducting block on the bias input that will decrease the inductive error signal by reducing the feedthrough of the drive signal to the output measurement.

References

[1] J. A. Brevik, N. E. Flowers-Jacobs, A. E. Fox, E. B. Golden, P. D. Dresselhaus and S. P. Benz, "Josephson arbitrary waveform synthesis with multilevel pulse biasing," *IEEE Trans. Appl. Supercond.*, vol. 27, no. 3, 1301707, April 2017.

[2] D. Zhao, H. E. van den Brom and E. Houtzager, "Mitigating voltage lead errors of an ac Josephson voltage standard by impedance matching," *Meas. Sci. Technol.*, vol. 28, 095004, March 2017.

Brevik, Justus; Donnelly, Christine; Flowers-Jacobs, Nathan; Fox, Anna; Hopkins, Peter; Dresselhaus, Paul; Benz, Samuel.
"Radio-frequency Waveform Synthesis with the Josephson Arbitrary Waveform Synthesizer."
Paper presented at 2019 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Measurement of Leakage Current to Ground in Programmable Josephson Voltage Standards

Alain Rüfenacht, Charles J. Burroughs, Paul D. Dresselhaus, and Samuel P. Benz

National Institute of Standards and Technology, Boulder, CO 80305, USA
alain.rufenacht@nist.gov

***Abstract* — The voltage error associated with the leakage current of programmable Josephson voltage standards (PJVS) is one of the largest contributions to the uncertainty in direct comparison of voltage standards. Due to the parallel biasing scheme of the PJVS and the resulting multiple leakage paths, the quantities "leakage resistance" and "leakage current" are not necessarily equivalent. A method to measure and model the leakage current to ground for a given PJVS output voltage is presented. By upgrading simple components of the NIST system, the leakage current to ground can be reduced from 250 pA to 50 pA at 10 V.**

***Index Terms* — Digital-analog conversion, Josephson arrays, standards, superconducting integrated circuits, voltage measurement.**

I. Introduction

A voltage standard that floats relative to Earth potential offers the flexibility to choose the grounding node of a measurement circuit. This feature is important both for a programmable voltage standard (**PJVS**) [1] used with Kibble balances [2] and more generally for the measurement of a voltage source that is, by construction, referenced to Earth ground (for example a Zener standard on line-power). The leakage resistance to ground of a PJVS system is not infinite. Because a PJVS array requires a current bias, it is always connected to its bias electronics. The bias electronics have an unavoidable leakage current to ground (LCG) error which must be evaluated and reported as a measurement uncertainty. This effect may constitute the largest contributor to the uncertainty budget of direct PJVS comparisons.

This paper explores the effects of the complex leakage current paths of PJVS systems with a simple model and method to measure the LCG for a given PJVS voltage and presents two different methods to derive the equivalent leakage resistance to ground (LRG) of the entire system. This detailed analysis of the various leakage current contributors has allowed us to reduce the LCG of our PJVS system by a factor five.

II. LCG and LRG Measurements

The determination of the individual contribution of each possible leakage path to the total LRG of a PJVS system is not an easy or practical option given 24 current bias leads (in the NIST system) [3]. The multiple parallel leakage paths are shown schematically in Fig. 1 as orange shaded areas surrounding the various components of the current bias electronics and wiring. However, we can measure the combined

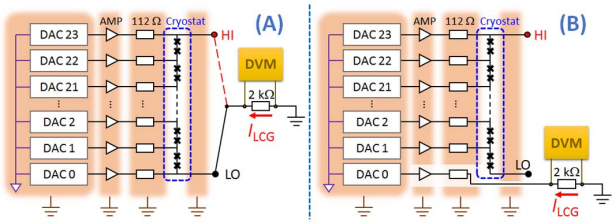


Fig. 1. Schematic of the LCG measurement (I_{LCG}) made by connecting the ground (A) to respectively the low side (LO) and the high side (HI) of the output voltage leads, and (B) to one of the DAC leads. The common node of the bias electronics is shown as a solid triangle symbol.

contribution of all the individual leakage currents to ground for a given PJVS voltage (I_{LCG}). After connecting either the low side or high side of the PJVS output leads to Earth ground potential (GND) through a 2 kΩ resistor (Fig.1A), I_{LCG} is determined by measuring the voltage drop across the resistor with a digital voltmeter (DVM) [4]. During the measurement acquisition, the voltage polarity of the PJVS is switched back and forth and the DVM acquisition starts after a delay of about 1 minute to remove both the influence of the thermal electromotive forces and the dielectric absorption. This method is highly sensitive to perturbations (electrostatic fields and microphonic noise) and this limits the overall relative accuracy to a few percent. The results of our LCG measurements versus PJVS voltage are shown in Fig. 2. Black circles and red squares represent the LCG measured when connecting the GND respectively to the low side (LO) and to the high side (HI) of the PJVS output leads. The measured LCG dependence is not entirely linear with the PJVS output voltage because the potential at the various bias nodes of the array do not necessarily increase linearly with the output voltage due to the ternary nature of the PJVS bias.

Another factor influencing the LCG is the potential difference ΔV between the common node (CMN) of the bias electronics and GND. The potential at the bottom node of the array is normally fixed at 0 V with respect to CMN. However, during this measurement the potential at the low side of the PJVS is adjusted first to −1 V and then −2 V with respect to CMN to produce voltages above respectively 8.5 V and 9.5 V. This is required due to the ±10 V dynamic range of the digital-to-analog converter (DAC). Grounding the high side of the array produces a larger LCG, mainly due to increased value of ΔV and the leakage path associated with the CMN node.

Rufenacht, Alain; Burroughs, Charles; Dresselhaus, Paul; Benz, Samuel.
"Measuring the Leakage Current to Ground in Programmable Josephson Voltage Standards."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

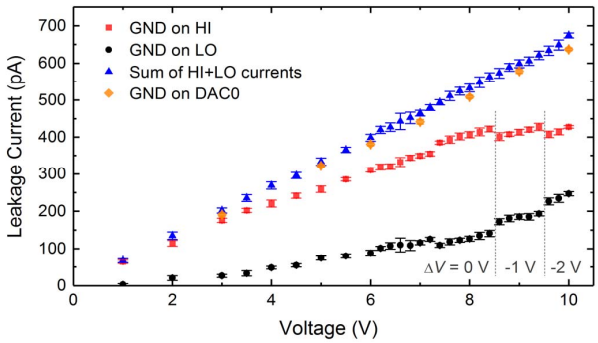


Fig. 2. Measured LCG as a function of the PJVS voltage for the two grounding configurations associated with the method A. Measured LCG as a function of the DAC0 voltage for the method B (orange diamonds).

The two independent measurements of LCG provide different values of the equivalent LRG (Table 1). However, these two measurements are complementary and by combining them (sum of the HI and LO LCG for a given PJVS voltage), each node of the system is virtually biased at the same voltage. As a result, the voltage dependence of the current sum (blue triangles) is now linear, as expected for a true leakage resistance measurement.

A second method, shown in Fig. 1B, allows a direct measurement of the LRG with a single measurement. One of the 24 DAC lines is disconnected from the PJVS array and connected through the sense resistor to ground (DAC0 line). Applying a voltage only on the DAC0 line elevates the potential difference of the whole system with respect to GND by the opposite voltage. The orange triangle symbols in Fig. 2 show the measured LCG as a function of the DAC0 voltage. The voltage dependence is linear and the LRG is comparable to the summed currents measured with the first method. Note that the two different LRG measurements were performed at different times.

PJVS Voltage	GND on	ΔV	Leakage current I_{LCG}	Equivalent Leakage Resistance
10 V	HI	+8 V	■ 427 pA	23 GΩ
10 V	LO	-2 V	● 246 pA	41 GΩ
10 V	SUM	(10 V)	▲ 673 pA	14.9 GΩ

Table 1. Measured LCG and LRG at 10 V.

III. Reduction of the LCG

After measuring the LRG of various components of the NIST system, we found that the LCG can be drastically reduced by replacing the commercial bias cables connecting the DAC to the amplifier board and the cable connecting the cryoprobe to the amplifier board with custom-made Teflon isolated cables. Figure 3 shows the LRG measurement after the cable replacement. No modification was applied to the commercially available DAC cards, the amplifier board (AMP), or the cryostat. The LCG at 10 V for the GND on LO configuration is

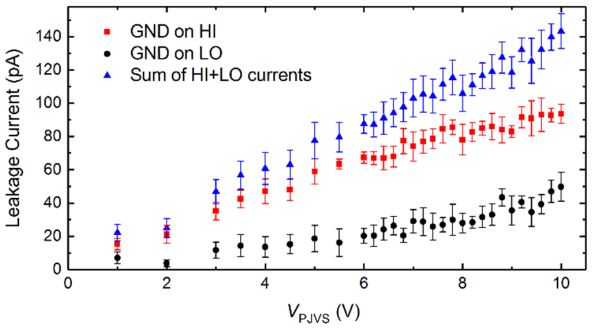


Fig 3. Measured LCG as a function of the PJVS voltage for the two grounding configurations associated with the method A, after replacement of the commercial bias cables with low leakage cables.

reduced by a factor five to around 50 pA. For a typical PJVS comparison at 10 V (assuming the GND is connected to the LO side of the opposite PJVS array and 2 Ω output lead resistance connecting the two systems), the 0.1 nV (or 1 part in in 10^{11}) voltage error due to such LCG is smaller than the noise floor of the digital nanovoltmeters currently used for such typical comparison measurement.

IV. Conclusion

Leakage currents to ground in the PJVS are often ignored or poorly estimated. An understanding of the contributions of the different elements of the bias electronics to the leakage resistance to ground of the entire system is the first step to improving the error due to leakage to ground. For metrology applications, the important quantity to measure is the leakage current to ground and its influence on the measurement circuit. The leakage current cannot be derived directly from the leakage resistance and must be measured independently for each PJVS output voltage. However, the leakage resistance to ground is a good indicator of the worst-case scenario; we are currently developing tools to automatically measure this quantity. To be compressive, the leakage current to ground from the bias electronics is not the only leakage effect: the leakage current due to the isolation resistance of the output leads must also be accounted for.

References

[1] C. A. Hamilton, C. J. Burroughs, and R. L. Kautz, "Josephson D/A converter with fundamental accuracy," *IEEE Trans. Instrum. Meas.,* vol. 44, no. 2, pp. 223–225, Apr. 1995.

[2] I. A. Robinson and S. Schlamminger, "The watt or Kibble balance: A technique for implementing the new SI definition of the unit of mass," *Metrologia,* vol. 53, no. 5, pp. A46–A74, 2016.

[3] C.J. Burroughs, P.D. Dresselhaus, A. Rufenacht A, D. Olaya, M.M. Elsbury, Y. Tang, and S.P. Benz, "NIST 10 V programmable Josephson voltage standard system," *IEEE Trans. Instrum. Meas.,* vol. 60, pp. 2482–2488, 2011.

[4] S. Solve, A. Rüfenacht, C. J. Burroughs and S. P. Benz, "Direct comparison of two NIST PJVS systems at 10 V," *Metrologia,* vol. 50, no. 5, pp. 441–451, 2013.

Rufenacht, Alain; Burroughs, Charles; Dresselhaus, Paul; Benz, Samuel.
"Measuring the Leakage Current to Ground in Programmable Josephson Voltage Standards."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)

Irena Bojanova
SSD, ITL
NIST
Gaithersburg, MD, USA
irena.bojanova@nist.gov

Yaacov Yesha
SSD, ITL; CSEE
NIST; UMBC
Gaithersburg; Baltimore MD, USA
yaacov.yesha@nist.gov

Paul E. Black
SSD, ITL
NIST
Gaithersburg, MD, USA
paul.black@nist.gov

***Abstract*—Random number generators may have weaknesses (bugs) and the applications using them may become vulnerable to attacks. Formalization of randomness bugs would help researchers and practitioners identify them and avoid security failures. The Bugs Framework (BF) comprises rigorous definitions and (static) attributes of bug classes, along with their related dynamic properties, such as proximate and secondary causes, consequences and sites. This paper presents two new BF classes: True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN). We analyze particular vulnerabilities and use these classes to provide clear BF descriptions. Finally, we discuss the lessons learned towards creating new BF classes.**

Keywords—randomness, random numbers, random number generators, pseudo-random number generators, software weaknesses, bug taxonomy, attacks.

I. Introduction

Randomness has application in many fields, including cryptography, simulation, statistics, politics, science, and gaming. Any specific use has its own requirements for randomness – e.g., random bit generation for cryptography or security purposes has stronger requirements than generation for other purposes. For cryptography or security purposes, the National Institute of Standards and Technology (NIST) recommends use of cryptographically secure Pseudo-Random Bit Generators (PRBGs). They are subject to the requirements in NIST SP 800-90A [8], NIST SP 800-90B [9] and NIST SP 800-90C [10]. Satisfying the requirements for a particular use can be surprisingly difficult [1] *.

Weaknesses (bugs) in random number generators (RNGs) may lead to wrong results from the algorithms that use the generated numbers or allow attackers to recover secret values, such as passwords and cryptographic keys. Formalization of randomness bugs would help researchers and practitioners identify them and avoid security failures. For that we have developed a general descriptive model of randomness and two randomness classes as part of the Bugs Framework (BF) [2, 3].

In this paper, we discuss randomness bugs, present the BF randomness bugs model, and detail our newly-developed randomness classes: True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN). The details include definitions and taxonomy of these classes, examples of vulnerabilities from the Common Vulnerabilities and Exposures (CVE) [4], and corresponding Common Weakness Enumeration (CWE) [5] or Software Fault Patterns (SFP) [6]. In the concluding section we discuss the lessons learned.

II. The Bugs Framework (BF)

The Bugs Framework (BF) provides accurate, precise, and unambiguous definitions of software weaknesses (bugs) and a language-independent taxonomy that allows clear descriptions of software vulnerabilities [2, 3]. It organizes bugs into distinct classes. The taxonomy of each BF class comprises: level, causes, attributes, consequences, and sites of bugs. Level (high or low) identifies the fault as language-related or semantic. Causes bring about the fault. At least one attribute (denoted as underlined) identifies the software fault, while the rest may be simply descriptive. It is useful to catalog possible consequences of faults. Sites are locations in code (identifiable mainly for low level classes) where the bug might occur under circumstances indicated by the causes.

Previously developed BF classes are: Buffer Overflow (BOF), Injection (INJ), Control of Interaction Frequency Bugs (CIF) [2], Encryption Bugs (ENC), Verification Bugs (VRF), Key Management Bugs (KMN) [3], and Faulty Result (FRS). Here we only give their definitions. Details and application examples of use are available at [7] .

BOF: *The software accesses through an array a memory location that is outside the boundaries of that array.*

INJ: *Due to input with language-specific special elements, the software assembles a command string that is parsed into an invalid construct.*

CIF: *The software does not properly limit the number of repeating interactions per specified unit.*

ENC: *The software does not properly transform sensitive data (plaintext) into unintelligible form (ciphertext) using a cryptographic algorithm and key(s).*

VRF: *The software does not properly sign data, check and prove source, or assure data is not altered.*

KMN: *The software does not properly generate, store, distribute, use, or destroy cryptographic keys and other keying material.*

FRS: *The software produces a faulty result due to conversions between primitive types, range violations, or domain violations.*

* The icon is used through the paper where we note the NIST SP 800-90 recommendations for construction of RBGs.

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

III. Randomness Bugs

A. Randomness Generation

We separate randomness generation in two distinct processes: true-random number generation and pseudo-random number generation. The former is nondeterministic true-randomness generation (full entropy), while the latter is deterministic pseudo-randomness generation.

True-random number generation uses entropy sources, while pseudo-random number generation uses true-random numbers as seeds. It is possible for a PRBG to use non-random seeds (e.g., for generating random numbers for simulation or game algorithms). PRBGs are used to extend the true-random seeds, produced from a True-Random Bit Generator (TRBG) output – if the seed has length `n`, the output of the PRBG can have length `m`, where `m>n`. However, a PRBG cannot increase the entropy of its seed.

Examples of randomness related attacks are direct RSA common factor attack [11, 12] cryptanalytic attack, input based attack, state compromise attack. [13]

B. The BF Randomness Model

Fig. 1 presents our BF randomness bugs model, showing in which software components of TRNG and Pseudo-Random Number Generator (PRNG) bugs can occur. It is a descriptive and not as a prescriptive model. It should not be used as a model for construction of Random Bit Generators (RBGs). (🔒 NIST SP 800-90C specifies construction of RBGs using the mechanisms and entropy sources described in SP 800-90A and SP 800-90B, respectively.) TRN is the name of our BF class of True-Random Number Bugs. PRN is the name of our BF class of Pseudo-Random Number Bugs. The BF randomness model helps identify where in the corresponding bugs could occur. TRN covers bugs related to entropy sources, TRBG, and TRNG. PRN covers bugs related to entropy pools, PRBG, and PRNG. Although, output from the former process may be used as input to the latter (see the red arrow in Fig. 1), they are distinct from the point of view that bugs related to each have different causes, attributes, and consequences. The random bits are optionally converted in a pseudo-random number based on the range that applications provide as an argument.

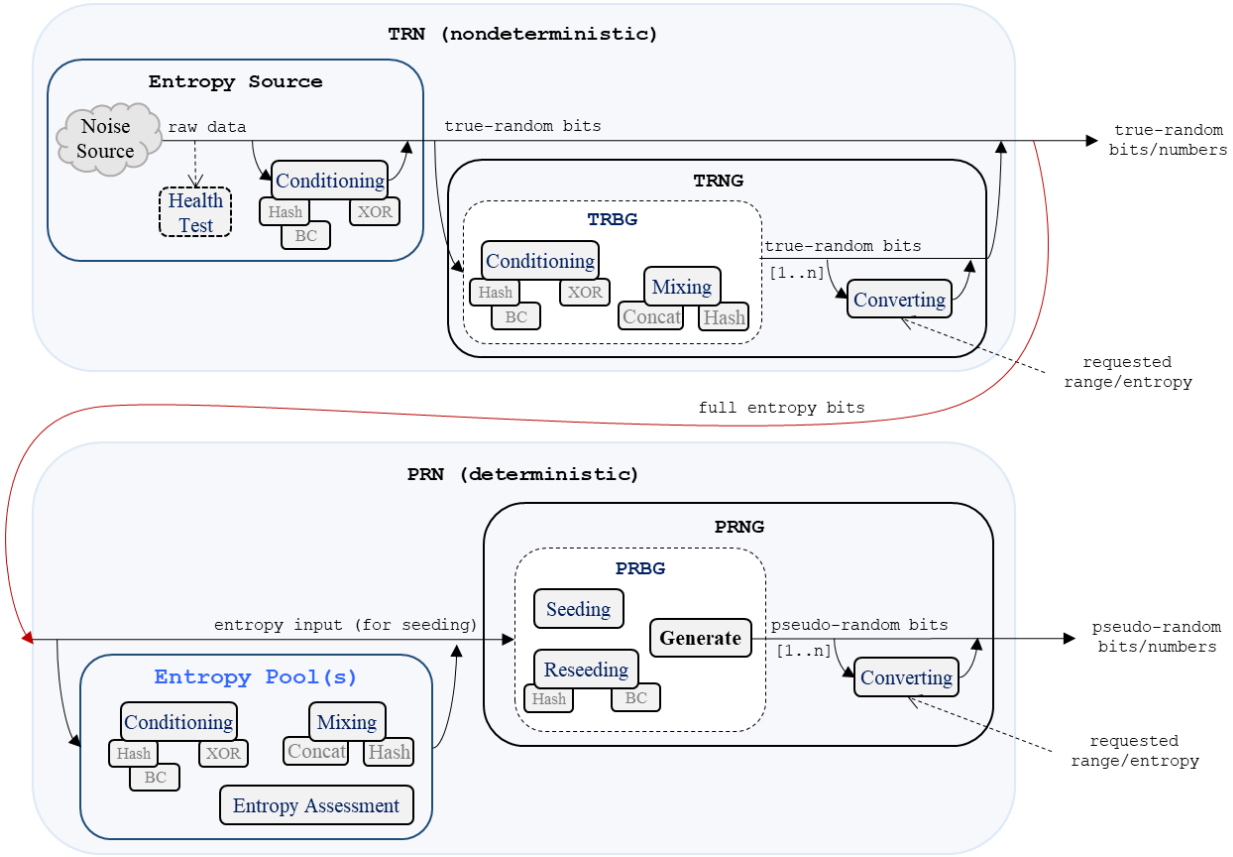


Fig. 1. The BF Model of Randomness Bugs. It is a descriptive and not as a prescriptive model. It should not be used as a model for construction of RBGs.
🔒 NIST SP 800-90A/B/C give specific requirements and architectures for appovoed RBGs.

(TRN – True-Random Number Bugs
PRN – Pseudo-Random Number Bugs
TRBG: True-Random Bit Generator
TRNG: – True-Random Number Generator
PRBG – Pseudo-Random Bit Generator
PRNG – Pseudo-Random Number Generator
BC – Block Cipher)

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

If live entropy source is used, the PRBG is said to support prediction resistance. A PRBG without prediction resistance can still be used where an on-demand entropy source and immediate resetting are not required. [8]

PRNGs are algorithmic and can have bugs. Most PRNGs are not cryptographically secure.

IV. True-Random Number Bugs Class - TRN

A. Definition

We define True-Random Number Bugs (TRN) as:

The software generated output does not satisfy all use-specific true-randomness requirements.

Note that the output sequence is of random bits, where values are obtained from one or more sources of entropy.

TRN is related to: PRN, ENC, VRF, KMN, IEX (Information Exposure Bugs).

B. Taxonomy

Fig. 2 depicts TRN causes, attributes and consequences.

The attributes of TRN are:

Function – Health Test, Conditioning, Mixing, Output, Converting.

Algorithm – Hash Function, Block Cipher, XOR, etc.

Used For – Seeding, Generation.

This is what the output sequence is used for. It could be used as a seed for a PRNG or for generation of passwords or cryptographic keying material (keys, nonces) [15].

Randomness Requirement – Sufficient Entropy, Sufficient Space Size, Non-Inferable. This is the failed requirement.

The importance of sufficient entropy is discussed in (CWE-333 [5]). The notion of entropy used in this paper is min-entropy, as the negative logarithm of the probability of the most likely outcome. Let `X` be a random variable such that the set of possible values that it can have is finite. Let `P` be the set of probabilities of `X` having those values. The min-entropy of `X` is defined as $-\log_2 \max(P)$, where `max()` finds the largest value in a set. The min-entropy is a measure of how difficult it is to guess the most likely entropy source output. [9, 16]

Space size is the number of elements of the space of possible outputs (CWE-334 [5]). If the number of different outputs is not sufficiently large, there is a vulnerability to a brute force attack. Non-inferable means one cannot recover from known (guessed) information anything about the TRBG output (CVE-2008-0141 [4]). TRBGs used for cryptography/security must satisfy the Non-Inferable randomness requirement.

In the graph of causes: Incorrect Entropy Assessment may result in TRN output having insufficient entropy. For example, a certain number of bits with full entropy may be needed, and because of Incorrect Entropy Assessment, the output may have insufficient entropy.

In the graph of consequences: Inadequate Input to PRNG could be repeating, weak, insufficiently random, predictable, small space seed or other input. "A program may crash or block if it runs out of random numbers" (CWE-333 [5]. Denial of Service (DoS) can be a direct consequence (CWE-333 [5]).

KMN could be a consequence as in finding private keys from public keys using a common factor attack [11, 12]. VRF could be a consequence of using a predictable random number with a signature algorithm, such as DSA (Digital Signature Algorithm). The information leaked (IEX) could be of the exact value of a generated random number (CWE-342 [5]) or a small range of values (CWE-343 [5]) from which the generated random number is easy to figure out.

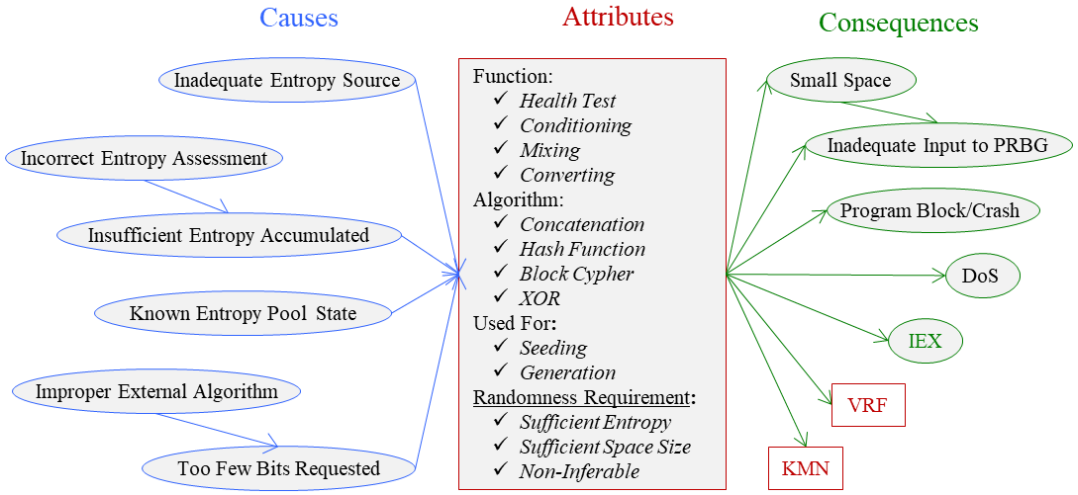


Fig. 2. The True-Random Number Bugs (TRN) represented as causes, attributes and consequences.

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

C. An Example

CVE-2008-0141

This vulnerability is listed in CVE-2008-0141 [4] and discussed in [17].

The BF TRN taxonomy for this vulnerability is:

Cause: Inadequate Entropy Sources (current date/time and user name)

Attributes:

Function: Mixing

Algorithm: Concatenation

Used for: Generation (of password)

Randomness Requirement: Non-Inferable (time known from password reset time, name - from user register)

Consequences: IEX (of password), leading to ATN (Authentication Fault)

Our BF description is:

Inadequate entropy sources (date/time and user name) mixing using concatenation allow generation of passwords that do not satisfy the non-inferable randomness requirement (time known from password reset time, name - from user register), which may be exploited for IEX (of password), leading to ATN.

Analysis (based on CVE-2008-0141 [4] and [17]): [17] includes "… The new password is simply the date (minute+seconds) and the var `$user` taken through `register_globals` (we can let it [be] empty) …". An attack exploiting that is described there.

D. Related CWEs and SFP

The related SFP cluster is SFP Primary Cluster: Predictability [6], which is CWE-905 with members CWE 330 to 344 [5].

Among them, the TRN CWEs are: 330, 331, 332, 333, 334, 337, 339, 340, 341, 342, 343 [5].

V. Pseudo-Random Number Bugs Class -- PRN

A. Definition

We define Pseudo-Random Number Bugs (PRN) as:

The software generated output does not satisfy all use-specific pseudo-randomness requirements.

Note that the output sequence is of random bits or numbers from a PRNG.

PRN is related to: TRN, ENC, VRF, KMN, IEX.

B. Taxonomy

Fig. 3 depicts PRN causes, attributes and consequences.

The attributes of PRN are:

Function – Conditioning, Mixing, Entropy Assessment, Seeding, Reseeding, Generate, Converting.

Algorithm – Concatenation, Hash Function, Block Cipher, XOR. Concatenation is the usual mixing of output from IID sources.

Used For – ASLR (Address Space Layout Randomization), Generation, Initialization, Input to Algorithm. This is what the output sequence is used for. It could be used for ASLR, generation of passwords or cryptographic keying material (keys, nonces) [15], initialization of cryptographic primitives [18] (e.g., an initialization vector for cipher block chaining mode of encryption; or a salt for hashing), or input to simulation, statistics, mathematics (e.g., Monte Carlo integration), or general algorithms.

Pseudo-Randomness Requirement – Unpredictability/ Indistinguishability, Prediction/Backtracking Resistance, Sufficient Space Size, Use Specific Statistical Tests. This is the failed requirement.

The pseudo-random output sequence should be statistically independent and unbiased [10]. It should pass the use-specific statistical tests for randomness. Unpredictability means that it should not be possible to predict next generated output from previous output [15]. Prediction Resistance means it is not possible to predict future output bits even if past or present state is known. Prediction resistance is not possible without a live entropy source. Backtracking Resistance means it is not possible to recover (backtrack) past output bits based on knowledge of the state at a given point in time [8, 10].

For Space Size see section IV.B. above. Indistinguishability for a PRNG means that its output is computationally indistinguishable (i.e., by any probabilistic polynomial time algorithm) from a truly random sequence [15].

PRNGs used for cryptography/security must satisfy the Unpredictability/Indistinguishability, Backtracking Resistance, and Sufficient Space Size requirements. Prediction Resistance however is not always required – e.g. for PIV cards.

C. Examples

1) *CVE-2001-1141*

This vulnerability is listed in (CVE-2001-1141 [4]) and discussed in [19, 20].

The BF PRN taxonomy for this vulnerability is:

Cause: Improper PRNG Algorithm (C `md_rand` – the secret PRNG state is updated with portion, as small as one byte, of the PRNG's previous output, which is not secret)

Attributes:

Function: Mixing (back into entropy pool)

Algorithm: Hash Function (SHA-1 – used for PRNG output and to update its internal secret state)

Used For: Generation (of cryptographic keying material – nonces, cryptographic keys)

Pseudo-Randomness Requirements: Sufficient Space Size and Unpredictability (can be predicted from previous value through brute force)

Consequences: KMN>Generate with IEX of future keying

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

Our BF description is:

Use of improper PRNG algorithm (C `md_rand` uses SHA-1 for mixing back in the entropy pool portion, as small as one byte, of previous output to update PRNG's state), allows generation of cryptographic keying material (nonces and cryptographic keys) that does not satisfy the sufficient space size and unpredictability (can be predicted from previous values through brute force) pseudo-randomness requirements, which leads to KMN>Generate and IEX of future keying material.

Analysis (based on CVE-2001-1141 [4] and [19 – 22]):

A PRNG used for cryptography does not satisfy the requirement of unpredictability from previous values, because the internal state can be determined from number of output requests. Possible consequences include: IEX of future PRNG output (CVE-2001-1141 [4]) (which is KMN>Generation failure) and weak encryption, confidentiality compromise [21] (which is ENC>Confidentiality failure [3]).

The entropy accumulation implementation (entropy pool and associated mixing function) allows reconstruction of the PRNG internal state [20]. The mixing hash function for `md` (in the C `md_rand`) gets half of the previous value of `md` and bytes from the PRNG internal state. Wrongly, the half used is the one with the PRNG's previous output (failed implementation relative to specification). Also, the number of used state bytes depends on the number of bytes requested as output, which could be as small as one byte. This enables a brute-force attack. The PRNG state could be reconstructed from the output of one large PRNG request (large enough to gain knowledge on `md`) followed by consecutive 1-byte PRNG requests [21, 22].

2) CVE-2008-4107

This vulnerability is listed in (CVE-2008-4107 [4]) and discussed in [23-26].

The BF PRN taxonomy for this vulnerability is

Cause: Improper PRNG Algorithms (not cryptographically strong PHP 5 `rand` and `mt_rand`)

Attributes:

Function: Generate (pseudo-random numbers)

Algorithms: e.g., LCG or LFSR, Mersenne Twister

Used For: Generation (of passwords)

Pseudo-Randomness Requirements: Unpredictability/ Indistinguishability and Prediction Resistance

Consequence: IEX (of password), leading to ATN

Our BF description is:

Improper PRNG algorithms (not cryptographically strong PHP 5 `rand` and `mt_rand`, based on algorithms such as LCG or LFSR, and Mersenne Twister) used to generate pseudo-random numbers, allow generation of passwords that do not satisfy the unpredictability/ indistinguishability and prediction resistance pseudo-randomness requirements and may be exploited for IEX of password, leading to ATN.

Analysis: PHP's `rand()` [23] usually uses LCG or LFSR, which is weak. PHP 5 `mt_rand()` [24] uses Mersenne Twister, which is weak as well. It enables finding the internal state and all future values from 624 values. [25]. This vulnerability can be used for password guessing (CVE-2008-4107 [4]), [26]).

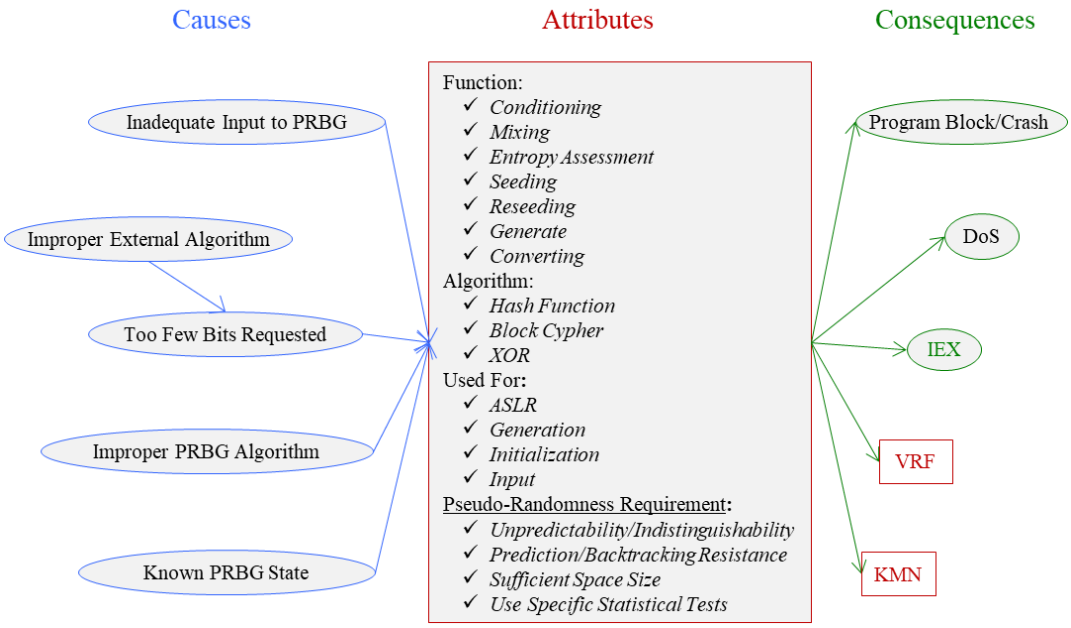


Fig. 3. The Pseudo-Random Number Bugs (PRN) represented as causes, attributes and consequences (ASLR – Address Space Layout Randomization).

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

CVE-2009-3238

This vulnerability is listed in (CVE-2009-3238 [4]) and discussed in [27-33].

The BF TRN and PRN taxonomy of this vulnerability is:

An TRN leads to a PRN.

TRN

Cause: Improper RNG Algorithm (same area is used for the `hash` array, allowing to repeatedly start from the same seed)

Attributes:

Functions: Mixing (of pid and jiffies), Conditioning

Algorithms: Concatenation, Hash Function (MD4)

Used For: Seeding

Randomness Requirements: Sufficient Entropy

Consequence: Inadequate Input to PRNG (repeating seed)

PRN

Cause: Inadequate Input to PRNG (repeating seed)

Attributes:

Function: Generate (pseudo-random numbers)

Used For: ASLR

Pseudo-Randomness Requirement: Unpredictability

Consequences: IEX of addresses

Our BF description is:

An TRN leads to a PRN.

TRN: *Improper RNG algorithm (same area is used for the `hash` array, allowing to repeatedly start from the same seed) for mixing (concatenation of pid and jiffies) followed by conditioning (using MD4 hash function), allows seeding that does not satisfy the sufficient entropy randomness requirement, leading to Inadequate Input to PRNG (repeating seed).*

PRN: *Inadequate input to PRNG (repeating seed), used to generate pseudo-random numbers, allows use of ASLR that does not satisfy the unpredictability pseudo-randomness requirement and may be exploited for IEX of addresses.*

Analysis (based on CVE-2009-3238 [4] and in [27-33]):

There is always a specific area for the `hash` array, allowing an attacker to repeatedly start from the same seed [32]. The result is predictability of address space randomization over a short time period. This violates the sufficient entropy requirement.

"Address space randomization hinders some types of security attacks by making it more difficult for an attacker to predict target addresses" [27]. CVE-2009-3238 [4] includes as a reference [33] that describes a fix that increases the randomness of ASLR: "Following security issues were fixed: ... CVE-2009-3238: The randomness of the ASLR methods used in the kernel was increased."

The TRN consequence is Inadequate Input to PRNG (repeating seed) [32]. The PRN consequence is IEX of addresses [27-30].

*3) **CVE-2017-15361*** (ROCA – Return Of the Coppersmith Attack)

This vulnerability is listed in CVE-2017-15361 [4] and discussed in [34].

The BF KMN [3] with inner PRN (see RND>Inadequate in [3]) taxonomy for this vulnerability is:

A KMN with inner PRN**.**

KMN:

Cause: Improper Algorithm Step (for generation of primes for RSA keys) leads to inner PRN

Attributes:

Cryptographic Data: Keying Material (keys)

Algorithm: RSA (key generation from two primes)

Operation: Generate (pair of public and private keys)

Consequences: Weak Public Key, which leads to IEX of Private Key.

Inner PRN:

Cause: Improper External Algorithm (generation of primes for RSA keys from random numbers and a constant related to keys size) leads to Too Few Bits Requested

Attributes:

Functions: Converting, Seeding (low entropy requested)

Used for: Generation (of secret prime numbers)

Randomness Requirement: Sufficient Space Size (e.g., one random number is only `37` bits for 512-bit RSA keys)

Consequence: IEX of generated primes (which format allows keys fingerprinting, factorization with Coppersmith algorithm, and finding random numbers and primes).

Our BF description is:

*A KMN with inner PRN**.***

KMN: *Improper algorithm step (for generation of primes for RSA keys) leads to inner PRN>Inadequate and allows generation of keying material (pair of public and private keys) with weak public key, leading to IEX of the private key.*

Inner PRN: *Improper external algorithm (generation of primes for RSA keys, from random numbers and a constant related to keys size) leads to too few bits requested at converting and at seeding, and allows generation of random numbers not satisfying the sufficient space size requirement*, which may be exploited for IEX of primes through fingerprinting of keys and factorization with Coppersmith algorithm.

Analysis (based on CVE-2017-15361 [4] and [34]): The generated RSA primes have the form: `p = k * M + (65537^a mod M)`. Where, `k` and `a` are random numbers; `M`

is a primorial (the product of the first `n` successive primes), related to the key size, which is a multiple of 32 bits. For keys with size in the `[512; 960]` interval, `n = 39` is used (i.e., `M = 2 * 3 * … * 167`); `n = 71; 126; 225` is used for key sizes within intervals `[992; 1952]`, `[1984; 3936]`, `[3968; 4096]`, respectively.

For keys with the same size, the generated RSA primes differ only in the values of `k` and `a`. The size of `M` is large – almost comparable to the size of the prime `p` (e.g., `M` has `219` bits for the 256-bit prime `p` used for 512-bit RSA keys). Since `M` is large, the sizes of `k` and `a` are small (e.g., `k` has `256 - 219 = 37` bits and `a` has `62` bits for 512-bit RSA). Hence, the resulting RSA primes suffer from a significant loss of entropy (e.g., a prime used in 512-bit RSA has only `99` bits of entropy), and the pool from which primes are randomly generated is reduced (e.g., from `2^256` to `2^99` for 512-bit RSA). [34]

The specific format of the primes allows fingerprinting of the keys followed by factorization. The size of `M` is large, but the logarithm $\log_{65537} N \bmod M$ can be computed, as `M` has only small factors. Factoring `N` using the Coppersmith's algorithm reveals `a` and `k` and eventually the prime `p`. [34]

D. Related CWEs and SFP

The related SFP cluster is SFP Primary Cluster: Predictability [6], which is CWE-905 with members CWE 330 to 344 [5].

Among them, the PRN CWEs are: 330, 331, 332, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343 [5].

VI. Conclusion

A. Summary

In this paper, we presented two new BF Classes: True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN). They join other rigorously-defined BF classes, such as Encryption/Decryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN). We presented the (static) attributes of the classes, along with the classes' causes and consequences. We analyzed particular vulnerabilities related to those classes and provided clear descriptions. We showed that the BF-structured descriptions of randomness bugs are quite concise, while still far clearer than unstructured explanations that we have found.

B. Lessons Learned and Future Work

At first, we tried to define a single Randomness class. However, that meant mixing the causes, attributes, and consequences related to true-random number generation and pseudo-random number generation. While exploring related vulnerabilities, we also realized that some of the consequences from the former may be causes for the latter. For example, if the consequence of a faulty seed generation is weak seed, this becomes the cause for a PRNG fault.

We considered keeping it all in one class by grouping the causes, attributes' values, and consequences related to true- and pseudo-random number generation. However, the model of randomness generation that we developed showed that these are two clearly separated processes and there should be two separate PRN and PRN randomness classes.

Work on explaining more randomness bugs using TRN and PRN will help us determine if our BF taxonomy needs refinement.

Our goal is for BF to become software developers' and testers' "Best Friend."

VII. References

[1] D. Eastlake, J. Schiller, S. Crocker, "Randomness Requirements for Security", 2005, https://tools.ietf.org/html/rfc4086/.

[2] I. Bojanova, P. E. Black, Y. Yesha, and Y. Wu, "The Bugs Framework (BF): A Structured approach to express bugs", Proceedings of IEEE International Conference on Software Quality, Reliability and Security (QRS), 2016, , pp. 175-182.

[3] I. Bojanova, P. E. Black, and Y. Yesha, "Cryptography classes in Bugs Framework (BF): Encryption Bugs (ENC), Verification Bugs (VRF), and Key Management Bugs (KMN)", 28th Annual IEEE Software Technology Conference (STC), 2017.

[4] The MITRE Corporation, Common Vulnerabilities and Exposures (CVE), https://www.cve.mitre.org.

[5] The MITRE Corporation, Common Weakness Enumeration (CWE), https://cwe.mitre.org.

[6] N. Mansourov, "DoD Software Fault Patterns," KDM Analytics, Inc., 2011. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADB381215/.

[7] The Bugs Framework, https://samate.nist.gov/BF.

[8] E. Barker, J. Kelsey, "NIST Special Publication 800-90A, Revision 1, Recommendation for Random Number Generation Using Deterministic Random Bit Generators", NIST, June 20q5, http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-90Ar1.pdf.

[9] Meltem Sönmez Turan, Elaine Barker, John Kelsey, Kerry A. McKay, Mary L. Baish, Mike Boyle, NIST Special Publication 800-90B, Recommendation for the Entropy Sources Used for Random Bit Generation, NIST, January, 2018, http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-90B.pdf,.

[10] E. Barker, J. Kelsey, (Second Draft) NIST Special Publication 800-90C "Recommendation for Random Bit Generator (RBG) Constructions", https://csrc.nist.gov/csrc/media/publications/sp/800-90c/draft/documents/sp800_90c_second_draft.pdf.

[11] N. Heninger, "New research: There's no need to panic over factorable keys–just mind your Ps and Qs", Feb. 15, 2012, https://freedom-to-tinker.com/2012/02/15/new-research-theres-no-need-panic-over-factorable-keys-just-mind-your-ps-and-qs/.

[12] A. K. Lenstra, J. P. Hughes, M. Augier, J. W. Bos, T. Kleinjung, and C. Wachter, "Ron was wrong, Whit is right", Infscience, https://infoscience.epfl.ch/record/174943.

[13] J. Kelsey, B. Schneier, D. Wagner, C. Hall. "Cryptanalytic attacks on pseudorandom number generators". Fast Software Encryption, Fifth International Workshop Proceedings. Springer-Verlag. pp. 168–188, 1998.

[14] J. Kelsey, B. Schneier, N. Ferguson, "Yarrow-160: Notes on the Design and Analysis of the Yarrow Cryptographic Pseudorandom Number Generator", https://www.schneier.com/academic/paperfiles/paper-yarrow.pdf.

[15] Wikipedia, "Cryptographically secure pseudorandom number generator",https://en.wikipedia.org/wiki/Cryptographically_secure_pseudorandom_number_generator/ .

[16] John Kelsey, Kerry A. McKay, and Meltem Sönmez Turan, "Predictive Models for Min-Entropy Estimation", International Workshop on Cryptographic Hardware and Embedded Systems, CHES 2015, pp 373-392, http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=918415.

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

[17] WebPortal CMS 0.6-beta - Remote Password Change, Exploit Database, https://www.exploit-db.com/exploits/4835/.

[18] Wikipedia, "Initialization vector", .https://en.wikipedia.org/wiki/Initialization_vector.

[19] VULDB, "OPENSSL/SSLEAY Pseudo-random Number Generator Weak Encryption", https://vuldb.com/?id.16976.

[20] Security Focus, "OpenSSL PRNG Internal State Disclosure Vulnerability", https://www.securityfocus.com/bid/3004.

[21] OpenSSL, "Weakness of the OpenSSL PRNG in versions up to OpenSSL 0.9.6a", "https://www.openssl.org/news/secadv/prng.txt

[22] Security Traker, "OpenSSL Uses Potentially Predictable Pseudo-Random Number Generator", https://securitytracker.com/id/1001961

[23] PHP Manual, rand, http://php.net/manual/en/function.rand.php.

[24] PHP Manual, mt_rand, .http://php.net/manual/en/function.mt-rand.php.

[25] StackExchange, "Information Security, How insecure are PHP's rand functions?", https://security.stackexchange.com/questions/18033/how-insecure-are-phps-rand-functions .

[26] Security Focus, "WordPress Random Password Generation Insufficient Entropy Weakness", http://www.securityfocus.com/bid/31115 .

[27] Wikipedia, "Address space layout randomization", https://en.wikipedia.org/wiki/Address_space_layout_randomization.

[28] GitHub, Inc. [US], "torvalds/linux", https://github.com/torvalds/linux/blob/master/drivers/char/random.c.

[29] J. Salwan, "ASLR implementation in Linux Kernel 3.7", Jan. 19, 2013, http://shell-storm.org/blog/ASLR-implementation-in-Linux-Kernel-3.7.

[30] xorl.wordpress, "Linux kernel ASLR Implementation" with 4 comments, https://xorl.wordpress.com/2011/01/16/linux-kernel-aslr-implementation.

[31] git.kernel, index:: kernel/git/torvalds/linux.git", https://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git/commit/?id=8a0a9bd4db63bc45e3017bedeafbd88d0eb84d02.

[32] xorl.wordpress, "CVE-2009-3238: Linux kernel get_random_int() Predictable Numbers", https://xorl.wordpress.com/2009/10/26/cve-2009-3238-linux-kernel-get_random_int-predictable-numbers.

[33] openSUSE Mailinglist Archive: opensuse-security-announce (12 mails) https://lists.opensuse.org/opensuse-security-announce/2009-11/msg00005.html.

[34] M. Nemec, M. Sys, P. Svenda, D. Klinec, V Matyas, "The Return of Coppersmith's Attack: Practical Factorization of Widely Used RSA Moduli", ACM CCS 2017, https://crocs.fi.muni.cz/_media/public/papers/nemec_roca_ccs17_preprint.pdf.

Bojanova, Irena; Yesha, Yaacov; Black, Paul.
"Randomness Classes in Bugs Framework (BF): True-Random Number Bugs (TRN) and Pseudo-Random Number Bugs (PRN)."
Paper presented at 42nd annual IEEE Computer Society COMPSAC conference: Staying Smarter in a Smartening World, Tokyo, Japan. July 23, 2018 - July 27, 2018.

Three Volt Pulse-Driven Josephson Arbitrary Waveform Synthesizer

Nathan E. Flowers-Jacobs, Alain Rüfenacht, Anna E. Fox, Steven B. Waltman, Justus A. Brevik, Paul D. Dresselhaus, and Samuel P. Benz

National Institute of Standards and Technology, Boulder, CO 80305, USA
nathan.flowers-jacobs@nist.gov

***Abstract* — We describe a new generation of Josephson Arbitrary Waveform Synthesizers that generate programmable ac waveforms with an rms amplitude of 3 V and, at 1 kHz, have a quantum locking range greater than 1 mA. This system has two chips with a total of 204 960 Josephson junctions (JJs) co-located on a cryocooler. To test for systematic errors, we phase-shift by 180° the waveform generated by half the JJs (102 480 JJs) to produce an approximate null. The measured residual of 51 nV rms implies a relative agreement of 3 parts in 10^8 between the two halves of the system.**

***Index Terms* — Digital-analog conversion, Josephson junction arrays, measurement standards, signal synthesis, superconducting integrated circuits, voltage measurement.**

I. Introduction

Over the past two decades, there has been steady improvement in the magnitude of the output voltage of the pulse-biased ac Josephson voltage standard, also known as the Josephson Arbitrary Waveform Synthesizers (JAWS) [1]. These continued improvements have resulted in the JAWS becoming a more practical tool over a larger amplitude range for ac calibrations and useful for applications such as impedance metrology [2]. Increasing the JAWS output voltage also increases the signal-to-noise ratio in measurements of systematic errors. Finally, the system improvements that were implemented to produce larger voltages have also led to simplified lower voltage systems with better understood sources of error.

In this paper, we further increase the JAWS output voltage by 50 % and measure the system performance. We use a cryocooled system to generate a 1 kHz waveform with a 3 V rms amplitude that is quantum-accurate over a dc current offset "locking" range of over ±0.5 mA. In Section II we describe the new chip and simplified electronics. In Section III we show measurements of the quantum locking range. We also begin an investigation of systematic errors by comparing two halves of the system using a null measurement and demonstrate a relative agreement of 3 parts in 10^8.

II. JAWS System Characteristics

This 3 V JAWS system combines two newly designed chips on a 4.6 K cryocooler (Fig. 1) with new bias electronics. The new chip design builds on advances demonstrated in [1]. The new bias electronics are based on a commercial pulse generator introduced in [3] and custom battery-powered isolation amplifiers that provide low-frequency compensation currents.

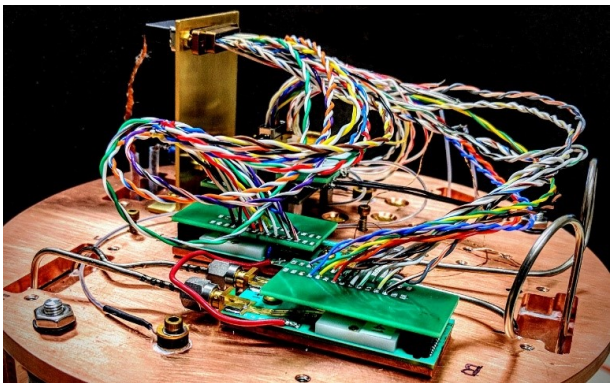


Fig. 1. Two JAWS chips mounted on a cryocooler.

The new chip design features eight arrays of Josephson junction (JJs); each array has 12 810 JJs that are embedded in series in the center conductor of a coplanar waveguide [1]. The JJs are self-shunted with a niobium-silicide barrier and niobium wiring. The JJ arrays have a critical current in liquid helium of 10 mA and a characteristic frequency of 20 GHz. In this paper, we demonstrate waveforms with an rms output of 1.5 V per chip.

Fast pulses are applied to the JJ arrays through two layers of Wilkinson dividers [1], so each chip requires only two RF inputs to drive all eight arrays. Each of the JJ arrays has an inside-outside DC block between the array and the Wilkinson dividers, with a 3 dB corner frequency of about 150 MHz to isolate the arrays at low-frequencies from each other and the RF input.

The fast pulses are provided by a commercial arbitrary waveform generator (Keysight M8195A*) at a rate of 14.4×10^9 pulses per second. An integrated finite impulse response filter is used to optimize the shape of the pulses so that every input current pulse causes each JJ to output exactly one quantized voltage pulse. The output of this generator is amplified with a commercial 50 GHz amplifier.

We use two different configurations for the low-frequency wiring of the chips. First, there is a "2 V" configuration where all the JJ arrays are connected in series with superconducting

*Commercial instruments are identified to specify the experimental procedure. NIST does not endorse commercial products. Other products may perform as well or better.

Flowers-Jacobs, Nathan; Rufenacht, Alain; Fox, Anna; Waltman, Steven; Brevik, Justus; Dresselhaus, Paul; Benz, Samuel.
"Three Volt Pulse-Driven Josephson Arbitrary Waveform Synthesizer."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

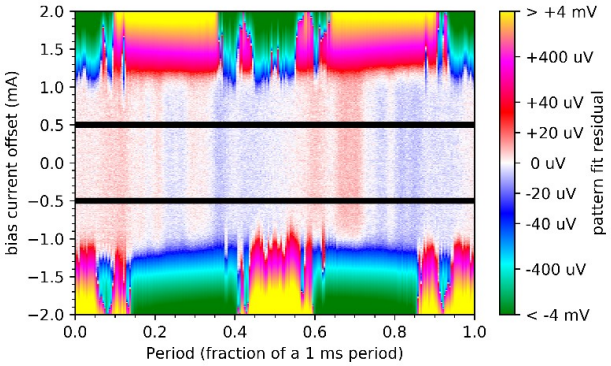


Fig. 2. Voltage residuals of a fit to a sinusoid of the digitally sampled waveform with an rms amplitude of 3 V versus dither offset current (**y**-axis) and waveform period (**x**-axis). The quantum locking range is greater than 1 mA (black lines).

wire. Second, in the "1 V + 1 V" configuration there are two pairs of outputs with the JJ arrays driven by each RF input separately connected in series. For this paper, we connect the two outputs of a "1 V + 1 V" chip and the output of a "2 V" chip in series on the cold stage of the cryocooler using short copper jumper wires.

As in earlier systems, we individually apply a low-frequency current to each JJ array through separate leads to compensate for the AC coupling of the fast input pulses by the inside-outside DC blocks [1]. The current to the arrays on each chip is provided by a pair of custom 4-channel battery-powered isolation amplifiers. The amplifier is driven by a function generator and the gain of each isolated output is controlled using an RS-485 interface on the amplifier.

For this experiment, we use one M8195A per chip. The two pulse generators are synchronized using a Keysight M8197A* synchronization module, and the instruments are synchronized to the same 10 MHz reference clock. The function generators are triggered from a M8195A marker channel.

III. 3 V and Null Measurements

We determine that we have successfully generated a 1 kHz waveform with a quantum-accurate rms magnitude of 3 V by measuring the output waveform as a function of dc bias current using a NI PXI-5922* digitizer. This measurement of the "quantum locking range" demonstrates that the system is operating correctly, with each JJ generating one output pulse for every input pulse, despite changes in the electrical or physical environment [1].

As is shown in Fig. 2, the quantum locking range is greater than 1 mA. During this measurement, the digitizer was set to its 10 V peak-to-peak range with a 1 MΩ input impedance and 1 MHz sampling rate. An additional factor-of-three resistive divider was inserted at the digitizer input (10 kΩ across the digitizer input in series with 20 kΩ) to reduce the effect of digitizer nonlinearities. The digitizer nonlinearities create

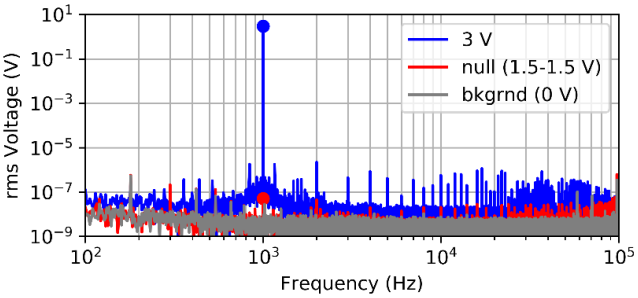


Fig. 3. Digitally sampled spectral measurement of a low distortion 1 kHz (dots) JAWS waveform with an rms magnitude of 3 V (blue), null measurement of 1.5 - 1.5 V (red), and 0 V background (grey).

vertical red/blue stripes that are independent of the offset current in Fig. 2 and create harmonics in the separately measured spectrum (Fig. 3).

As a first step in the investigation of systematic errors in this JAWS system, we measure the agreement between the different halves of the system. Each pulse generator is loaded with waveforms that are 180° out of phase. An additional delay is then applied in each pulse generator and between the pulse generators to minimize the combined amplitudes, resulting in an attempted null with an rms magnitude of 51 nV (Fig. 3). The magnitude of the attempted null depends on which JJ arrays and bias channels are generating in-phase/out-of-phase voltages. This is likely due to leakage current from the compensation electronics and will need to be investigated in more detail, along with other sources of systematic error.

VI. Conclusion

We demonstrate a cryocooled JAWS system that generates ac waveforms with an rms magnitude of 3 V and a quantum locking range at 1 kHz that is greater than 1 mA. A null measurement indicates that the two halves of the system agree to within 3 parts in 10^8, with each half generating a 1 kHz waveform with an rms magnitude of 1.5 V. More extensive measurements are in progress to better understand the systematic errors of this system.

References

[1] N. E. Flowers-Jacobs **et. al.**, "Josephson arbitrary waveform synthesizer with two layers of Wilkinson dividers and an FIR filter," **IEEE Trans. Appl. Supercond.,** vol. 26, no. 6, 143007, Sep 2016.

[2] F. Overney, N. E. Flowers-Jacobs, B. Jeanneret, A. Rüfenacht, A. E. Fox, J. M. Underwood, A. D. Koffman, and S. P. Benz, "Josephson-based full digital bridge for high-accuracy impedance comparisons," **Metrologia,** vol. 53, no. 4, pp. 1045-1053, Aug 2016.

[3] J. A. Brevik, N. E. Flowers-Jacobs, A. E. Fox, E. B. Golden, P. D. Dresselhaus, and S. P. Benz, "Josephson arbitrary waveform synthesis with multilevel pulse biasing," **IEEE Trans. Appl. Supercond.,** vol. 27, no. 3, 1301707, Apr 2017.

Flowers-Jacobs, Nathan; Rufenacht, Alain; Fox, Anna; Waltman, Steven; Brevik, Justus; Dresselhaus, Paul; Benz, Samuel.
"Three Volt Pulse-Driven Josephson Arbitrary Waveform Synthesizer."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Persistence of Electromagnetic Units in Magnetism

Ronald B. Goldfarb

National Institute of Standards and Technology, Boulder, CO 80305, USA
ron.goldfarb@nist.gov

***Abstract* — The centimeter-gram-second system of electromagnetic units (EMU) has been used in magnetism since the late 19th century. The International System of Units (SI), a successor to Giorgi's 1901 meter-kilogram-second system, was adopted by the General Conference on Weights and Measures in 1960. However, EMU remains in common use for the expression of magnetic data. The forthcoming revision of the SI provides an excellent reason for magnetics researchers to abandon the use of EMU.**

***Index Terms* — International System of Units, electromagnetic units, permeability of vacuum.**

I. Introduction

The International System of Units (SI), established in 1960 by the General Conference on Weights and Measures, has been generally accepted by researchers in most scientific disciplines. Magnetics is a notable exception, where the centimeter-gram-second (CGS) system of electromagnetic units (EMU), as formulated by Maxwell in 1873, remains in common use. The coexistence of SI and EMU in magnetics, and the conversion from one to the other, has been a source of confusion and error for students and practitioners.

Although the use of SI in magnetics has some inconveniences, they are minor compared to the ambiguities in the EMU system. The case for the SI remains compelling for the reasons first articulated by Giovanni Giorgi at the beginning of the 20th century. Importantly, the expected revision of the SI will make it incompatible with EMU.

II. Advantages and Disadvantages of SI and EMU

One of the advantages of SI is that it unifies magnetic and electrical units, whereas CGS bifurcates into EMU and electrostatic units (ESU). A possible disadvantage in SI is that two constitutive relations are recognized: $B = \mu_0(H + M)$, the Sommerfeld convention, where B is magnetic flux density, μ_0 is the permeability of vacuum, H is magnetic field strength, and M is the magnetization; and $B = \mu_0 H + J$, the Kennelly convention, where J is the magnetic polarization. Because M and J have different units, confusion is averted.

A disadvantage of SI is that the units for H, amperes per meter, are too small. Researchers have the urge to use tesla for H (which they could use if they instead wrote $\mu_0 H$), or they mistakenly refer to B instead of H.

Turning to EMU, the disadvantages are more serious. The unit for magnetic moment m is often expressed as "emu"; however, "emu" is not a unit, but is simply an indicator of electromagnetic units. The actual units for m are ergs per gauss or ergs per oersted.

As in SI, volume susceptibility is dimensionless, but its value in EMU is smaller than in SI. It may be appreciated that values of volume susceptibility, reported in the literature as dimensionless in both SI and EMU, might be difficult to compare. In EMU, volume susceptibility is often expressed in "emu," "emu per cubic centimeter," or "emu per cubic centimeter per oersted," a state of confusion originating from the misuse of "emu" as the unit for magnetic moment.

Magnetization (magnetic moment per unit volume) is commonly expressed either as M in "emu per cubic centimeter" or as $4\pi M$ in units of gauss. They are dimensionally equivalent, but they differ numerically by 4π. This double definition often leads to misunderstandings and mistakes.

Care is required when electrical and magnetic quantities are combined: the EMU unit of current is the abampere. Some researchers are reluctant to abandon the ampere and use mixed units in equations that do not balance dimensionally.

III. The Permeability of Vacuum in the Revised SI

The forthcoming revision of the SI [1], in which fixed values will be assigned to the Planck constant h and the elementary charge e [2], will accentuate the philosophical differences between EMU and SI: μ_0 is fixed at unity in the former but will be measurable, in principle, in the latter [3], and quantities will no longer be strictly convertible by factors of 4π and powers of 10 [4].

In the revised SI, μ_0 will be derived from fixed constants h, e, and the speed of light c, and the experimentally determined fine structure constant α [3]:

$$(\mu_0)_{\text{experimental}} = (2h/ce^2)_{\text{fixed}} \cdot (\alpha)_{\text{experimental}} . \qquad (1)$$

The value of μ_0, initially equal $4\pi \times 10^{-7}$ H/m to 9 significant figures, may change slightly over time as better measurements of α are made [3]. Magnetics researchers will have to choose to work and publish in one of two incompatible systems: EMU, which has a long tradition, or SI, which unifies all metrology and which has been adopted by international convention.

Of course, the adoption of SI by magneticians accustomed to working in EMU will require not only relatively straightforward conversions of units, but the conversion of equations (e.g., insertions and deletions of μ_0 and 4π).

Goldfarb, Ronald.
"Persistence of Electromagnetic Units in Magnetism."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Fortunately, it is much easier to verify the dimensional consistency of equations in SI than in EMU.

IV. Giorgi's Rationalized MKS System

The demotion of μ_0 from its immutable value of $4\pi \times 10^{-7}$ H/m within the SI might seem to violate the sanctity of a fixed constant. However, when a rationalized, four-dimensional system was first proposed by Giorgi [5], both μ_0 and the permittivity of vacuum ε_0 were regarded as subject to experimental refinement, with $\mu_0 \approx 1.256 \times 10^{-6}$ H/m and $\varepsilon_0 \approx 8.842 \times 10^{-12}$ F/m, "free from any unnecessary 4π," and both subject to the condition $(\mu_0 \varepsilon_0)^{-½} = c \approx 3 \times 10^8$ m/s [6].

That same condition applies in EMU, with $\mu_0 = 1$ and $\varepsilon_0 = c^{-2}$, and in ESU, with $\mu_0 = c^{-2}$ and $\varepsilon_0 = 1$. In 1905, Giorgi noted that his rationalized system "is neither electrostatic nor electromagnetic, because neither the electric nor the magnetic constant of free ether [vacuum] is assumed as a fundamental unit" [6].

V. Is B the Same as H in Vacuum?

One of the appealing aspects of EMU is that, in vacuum, B is equal to H in value and dimensions, despite their different unit names (gauss and oersted). Whether the fields B and H in vacuum are physically the same in EMU (and in the present SI) is historically controversial [7].

In the present SI, B and H in vacuum are mutually convertible through the fixed constant μ_0 despite their different values and dimensions. This, too, is appealing. It is similar to the conversion of B and H in ESU (if one assumes c is a fixed constant in CGS; no one really knows because no international standards organization maintains the CGS system).

In the revised SI, B and H will not be similarly convertible. In vacuum, one could, in principle, measure either B or H, depending on the experiment, and calculate its counterpart using the most recent value of μ_0. Or one could, in principle (if not in reality), measure both B and H and calculate a new value of μ_0. Thus, it seems implicit in the revised SI that B and H in vacuum are physically different.

Before the value of c was fixed in the SI in 1983, the same considerations applied to (1) the electric flux density (electric displacement) D and the electric field strength E in SI and (2) B and H in ESU: they were mutually convertible through ε_0, a constant whose value depended on the measured value of c.

VI. Conclusion

In introducing his rationalized meter-kilogram-second system in 1901 [5], which later evolved into the meter-kilogram-second-ohm [6] and then the meter-kilogram-second-ampere systems, Giorgi wrote, *"Il sistema CGS, con questo, perde ogni ragione di esistere; ma non credo che il suo abbandono sarà lamentato da alcuno."* ("With this, the CGS system loses every reason to exist; but I do not think that its abandonment will be lamented by anyone.")

He may be correct, eventually.

References

[1] CIPM Decision 106-10, October 2017: "The International Committee for Weights and Measures (CIPM) welcomed recommendations regarding the redefinition of the SI from its Consultative Committees. The CIPM noted that the agreed conditions for the redefinition are now met and decided to submit draft Resolution A to the 26th meeting of the General Conference on Weights and Measures (CGPM) and to undertake all other necessary steps to proceed with the planned redefinition of the kilogram, ampere, kelvin and mole." [Online] https://www.bipm.org/en/committees/cipm/meeting/106.html

[2] D. B. Newell, F. Cabiati, J. Fischer, K. Fujii, S. G. Karshenboim, H. S. Margolis, E. de Mirandés, P. J. Mohr, F. Nez, K. Pachucki, T. J. Quinn, B. N. Taylor, M. Wang, B. M. Wood, and Z. Zhang, "The CODATA 2017 values of h, e, k, and N_A for the revision of the SI," Metrologia, vol. 55, no. 1, pp. L13–L16, Feb 2018; doi: 10.1088/1681-7575/aa950a.

[3] R. B. Goldfarb, "The permeability of vacuum and the revised International System of Units," IEEE Magnetics Letters, vol. 8, 1110003, Dec 2017; doi: 10.1109/LMAG.2017.2777782.

[4] R. S. Davis, "Determining the value of the fine-structure constant from a current balance: Getting acquainted with some upcoming changes to the SI," American Journal of Physics, vol. 85, no. 5, pp. 364–368, May 2017.

[5] G. Giorgi, "Unità razionali di elettromagnetismo," Atti dell' Associazione Elettrotecnica Italiana, vol. 5, pp. 402–418, Oct 1901; reprinted as "Sul sistema di unità di misure elettromagnetiche," Il Nuovo Cimento, vol. 4, no. 1, pp. 11–30, Dec 1902; doi: 10.1007/BF02709460. An English translation appears in C. Egidi, ed., Giovanni Giorgi and His Contribution to Electrical Metrology, Politecnico di Torino, Turin, Italy, 1990, pp. 135–150.

[6] G. Giorgi, "Proposals concerning electrical and physical units," Journal of the Institution of Electrical Engineers, vol. 34, no. 170, pp. 181–185, Feb 1905; doi: 10.1049/jiee-1.1905.0013.

[7] J. J. Roche, "B and H, the intensity vectors of magnetism: A new approach to resolving a century-old controversy," American Journal of Physics, vol. 68, no. 5, pp. 438-449, May 2000; doi: 10.1119/1.19459.

Goldfarb, Ronald.
"Persistence of Electromagnetic Units in Magnetism."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

SIM Time Scale: 10 years of operation

J. M López R.[1,2], M. Lombardi[3], E. de Carlos L.[2], N. Diaz[2] and C. A. Ortiz C.[1,2]

[1]CINVESTAV, Libramiento Norponiente 2000, Querétaro, México
jm.lopez@cinvestav.mx

[2]CENAM, Carretera a los Cues km 4.5, Querétaro, México

[3]NIST, Boulder, Colorado, USA

***Abstract* — The *Inter-American Metrology System* (SIM) is one of the world´s five major Regional Metrology Organizations (RMO´s). Starting in 2005, the SIM Time and Frequency Metrology Working Group (SIM TFWG) developed a time and frequency comparison network for the Americas (SIMTN). Currently 23 NMIs from SIM region participate on the SIMTN. Since 2008 the SIM TFWG is producing a time scale called SIM Time Scale (SIMT). SIMT is an averaged time scale computed in near real-time (every hour) from time difference data that the SIMTN produces and publishes every 10 minutes. In this paper, a SIMT evaluation based on data from August 2016 to January 2018 is presented.**

***Index Terms* — Time Scales, international comparison, atomic clocks.**

I. Introduction

The measurement of time is of the utmost importance for many applications, including: global satellite navigation systems, communication networks, electric power transportation, astronomy, electronic transaction, and national defense and security. Both, the SIM Time Network (SIMTN) and the SIM Time Scale (SIMT) have led to better coordination and cooperation in the Americas in time and frequency metrology. The SIMT is believed to be the first multinational time scale whose results are computed and published in real time via the Internet, and 2018 will mark its 10-year anniversary. Within the SIM region, SIMT complements Coordinated Universal Time (UTC) by providing real-time support to operational timing and calibration systems. SIMT is sufficiently stable to measure the stability of most SIM local time scales and provides a good approximation of the UTC timing accuracy (±10 ns). We show in this paper how the reliability, accuracy, and stability of SIMT has improved during its 10 years of operation.

II. The SIMTN

The SIMTN became operational in 2005, and 23 nations have joined the network as of December 2017. The SIMTN continuously compare the time scales of all SIM local time scales with each other and produces measurement results in near real time [1]. The comparisons are performed via the global positioning system (GPS) common-view and all-in-view techniques with multichannel single-frequency (L1 band) receivers. SIMTN servers are located at National Research Council (NRC) in Canada, National Institute of Standards and Technology (NIST) in the USA and the Centro Nacional de Metrología (CENAM) in Mexico. The three SIMTN servers host identical software that processes and publishes measurement data whenever requested by a user (https://tf.nist.gov/sim).

III. The SIMT

The large number of local time scales at SIM region made it attractive to generate a composite time scale that could be distributed and shared. Work on SIMT began in early 2008 at CENAM and it has been refined through the years. SIMT is continuously operated and it is made publicly available in real time via the Internet. It includes local NMI time scales, SIM(k), as single clocks in the SIMT ensemble and it is not dependent on the time scale maintained by any individual national metrology institute (NMI). SIMT has been designed to be an instantly accessible reference standard that can be used to monitor the performance of the local SIM(k) scales and operational timing system in the short, medium and long terms.

The SIMT algorithm is published in [2]. It is designed to use exponential filtering to predict the time and frequency differences of SIMT(k) with respect to the averaged time scale. Clocks are weighted in terms of the inverse of their Allan deviation. The weighting criteria were modified in 2012 [3] to include an accuracy factor given by the inverse of the $|\langle \Delta f \rangle|$, where $|*|$ and $\langle * \rangle$ means absolute value and average value of $*$, respectfully. The Δf term is the frequency difference during the previous 240 hours between SIMT(k) and SIMT. Because SIMT is not steered to agree with UTC or any of the individual UTC(k) time scales, it can be considered to be a free running time scale.

IV. SIMT Performance

To evaluate the performance of the SIMT scale, we use the national time scale of Mexico, UTC(CNM), as a "common clock" to compute the time differences UTC – UTC(CNM), UTCr – UTC(CNM) and SIMT – UTC(CNM). The UTCr time scale is "rapid" UTC, a version of UTC that is published weekly with daily values [4], as opposed to UTC which is published monthly with values given at 5-d intervals. Figure 1 shows the

Lopez, J.; Lombardi, Michael; de Carlos, E.; Munoz, N.; Ortiz, C..
"SIM Time Scale: 10 years of operation."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

time differences of UTC(CNM) with respect to UTC, UTCr and SIMT scales from October 2015 to January 2018, showing that SIMT is in good agreement with both UTC and UTCr for medium and long-term intervals. However, in the short term it can be noticed that SIMT is less stable than UTC or UTCr. This is due to two main reasons. First, the number of clocks used to produce SIMT is significant smaller than the number of clocks used to compute UTC or UTCr scales. Second, the short-term noise in data produced with the SIMTN is larger than the noise in data used to calculate the UTC or UTCr. That is primarily because the SIM time transfer systems are single frequency (L1 band) Global Positioning System (GPS) receivers, as opposed to the multi-frequency GPS receivers and two-way satellite time transfer systems used by many NMIs to contribute to UTC and UTCr.

Figure 2 shows the time differences of UTC – SIMT, UTCr – SIMT and UTC – UTCr. Those time differences were obtained by utilizing UTC(CNM) as a "common clock". The outlier on the UTCr – SIMT graph near Modified Julian Date (MJD) 57750 is due to an erroneous UTCr published value.

Figure 3 shows the frequency instabilities of the UTC – UTC(CNM), SIMT – UTC(CNM), UTCr – UTC(CNM), UTC – SIMT, UTCr – SIMT and UTC – UTCr time differences. In the short and medium-term the three most stable comparisons are UTC – UTCr, UTC – SIMT and UTCr – UTC, in that order. On the other hand, the stabilities of the time differences of UTC(CNM) with respect to UTC, UTCr and SIMT are almost equivalent. This implies that SIMT, for the purposes of evaluating the frequency stability of UTC(CNM), is an equivalent reference to UTC and UTCr, and is more readily accessible.

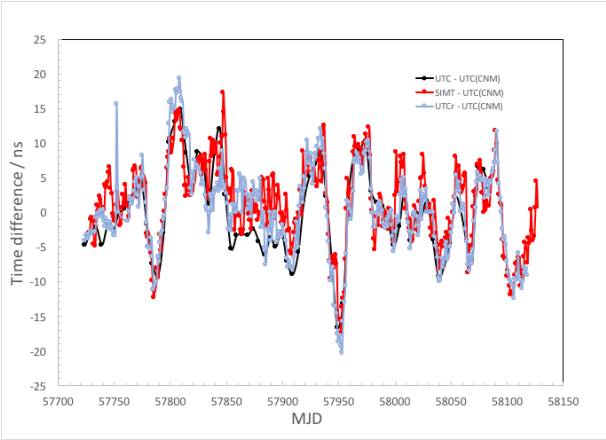


Figure 1. Time differences of the CENAM´s time scale UTC(CNM) with respect to the UTC, UTCr and SIMT scales from August 2016 to January 2018.

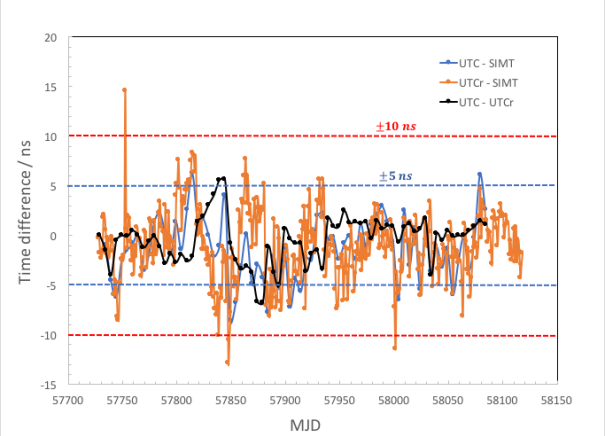


Figure 2. Time differences of UTC – SIMT, UTCr – SIMT and UTC – UTCr with UTC(CNM) as a "common clock".

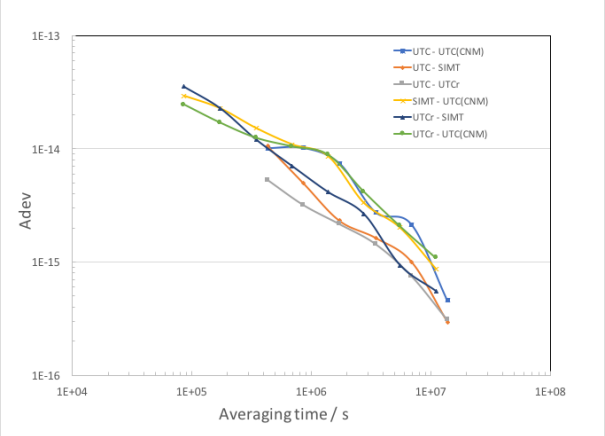


Figure 3. Frequency stabilities for the time differences UTC – UTC(CNM), SIMT – UTC(CNM), UTCr – UTC(CNM), UTC – SIMT, UTCr – SIMT and UTC – UTCr.

REFERENCES

[1] M. A. Lombardi, A. N. Novick, López-Romero, F. Jimenez, E. de Carlos-Lopez, J. S. Boulanger, R. Pelletier, R. de Carvalho, R. Solis, H. Sanchez, C.A. Quevedo, G. Pascoe, D. Perez, E. Bances, L. Trigo, V. Masi, H. Postigo, A. Questelles, and A. Gittens, "The SIM Time Network," *J. Res. Natl. Inst. Stan.*, vol. 116, no. 2, pp. 557-552, March-April 2011.

[2] J. M. López-Romero, M. A. Lombardi, N. Díaz-Muñoz, and E. de Carlos-Lopez, "SIM Time Scale," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 12, pp. 3343-3350, 2013.

[3] J. M. López-Romero, M. A. Lombardi, N. Díaz-Muñoz, and E. de Carlos-Lopez, "A New Algorithm for Clock Weights for the SIM Time Scale," *Proc. 2012 Simposio de Metrología*, Querétaro, México, October 2012.

[4] G. Petit, F. Arias, A. Harmegnies, G. Panfilo, and L. Tisserand, "UTCr: a rapid realization of UTC," *Metrologia*, vol. 51, no. 1, pp. 33-39, 2013.

Lopez, J.; Lombardi, Michael; de Carlos, E.; Munoz, N.; Ortiz, C..
"SIM Time Scale: 10 years of operation."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Improving packet synchronization in an NTP server

Andrew N. Novick[1], Michael A. Lombardi[1], Kevin Franzen[2], and John Clark[2]
[1]*Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado, USA*
[2]*Masterclock, St. Charles, Missouri, USA*
novick@nist.gov

ABSTRACT

A computer or dedicated client can use the Network Time Protocol (NTP) to synchronize an internal clock to a server that is synchronized by a 1 pulse-per-second (pps) signal from a national timing laboratory. Measuring an NTP server on a local area network can reveal timing synchronization errors and anomalies that are not nearly as likely to be recognized when NTP is utilized on a wide area network, where network delay asymmetry is the dominant source of uncertainty. We measured a commercially available NTP server by making rapid packet requests with UTC(NIST) as the common reference clock for both the server and client. Analysis of the results revealed repetitive synchronization errors in the packets transmitted by the server. Although these errors were usually too small to be detected by customers who deployed the server or by clients who accessed the server during typical usage, improving packet synchronization would benefit many applications. By collaborating with the server's manufacturer, the source of the problem was revealed, and firmware changes were made to remove the synchronization errors. This paper describes our measurements and how the results were used to improve the server's NTP packet synchronization.

1. INTRODUCTION

The Network Time Protocol (NTP) is the most common method for synchronizing computer clocks and devices over the public Internet. Time synchronization of a client within 1 ms of a national timing laboratory is possible across a country, or even internationally [1]. The most prevalent source of uncertainty across a packet-switched wide area network (WAN) is the asymmetry of the path delay between the client and server. There are situations where the server clock adds a noticeable offset to the result [2, 3], but because using NTP over a WAN is typically utilized to synchronize a clock to the nearest second, the server clock uncertainty is usually not discernible or of concern to the client. Also, NTP clients on the public Internet often use software that compares results from several NTP servers and discard information from servers with incorrect clocks.

The uncertainty of NTP on a local area network (LAN) is usually much smaller, primarily because the smaller round trip path delay limits the potential effects of network asymmetry, which cannot exceed 50% of the path delay. We have previously examined the limitations of using NTP on a LAN by connecting the client computer directly to the server and measuring server-client time differences of less than 1 µs on average over a 20-day period with a standard deviation less than 10 µs [4]. Adding network components such as routers and switches to reach a different subnet added several microseconds to the average and more than doubled the standard deviation, but the result was still much better than using a WAN. Also, in a comparison of multiple NTP servers with the same reference clock and connected to the same network, the results differed by tens of microseconds [5]. These examples show that when NTP is used on a LAN, the uncertainty of the packet synchronization provided by the server is a significant factor in the uncertainty of the time received by the client. This paper shows how we measured anomalies, outliers, and server offsets, and then shared the results with a server manufacturer, helping them to achieve better NTP packet synchronization.

2. DETECTION OF SERVER ANOMALIES

In a previous paper where we compared several NTP servers, our results showed unexplained time steps in data from two servers, made by the same manufacturer, that occurred at the same time [5]. The data from the other servers being measured on the same network did not indicate anything atypical. We began collaborating with the server manufacturer to investigate these results, and set up a test bed with multiple NTP servers using a 1 pulse per second (pps) signal from UTC(NIST) as their reference input.

Novick, Andrew; Lombardi, Michael; Franzen, Kevin; Clark, John.
"Improving packet synchronization in an NTP server."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) ,
Reston, VA, United States. January 29, 2018 - February 1, 2018.

The NTP servers were compared to UTC(NIST) with a client computer measurement system used in previous experiments [1, 4, 5]. Every 10 s, the client software sequentially requests packets from a list of NTP servers and compares the time stamp obtained in the packets to UTC(NIST). Several of the servers on the list are located on the same subnet. Because both the server and client computers are referenced to UTC(NIST), this is a direct test of the uncertainty of the packet synchronization of each server. Figure 1 shows the round trip delay and time difference of Server A for a period of four days. In the middle section of the graph, Server B was added to the same subnet, and it appeared to cause a time step of over 80 µs in the Server A data. The round-trip delay for Server A did not change when Server B was added, so we did not believe that the network path was affected. The servers shared a network hub and it was initially predicted that crosstalk from Server B caused the change in Server A.

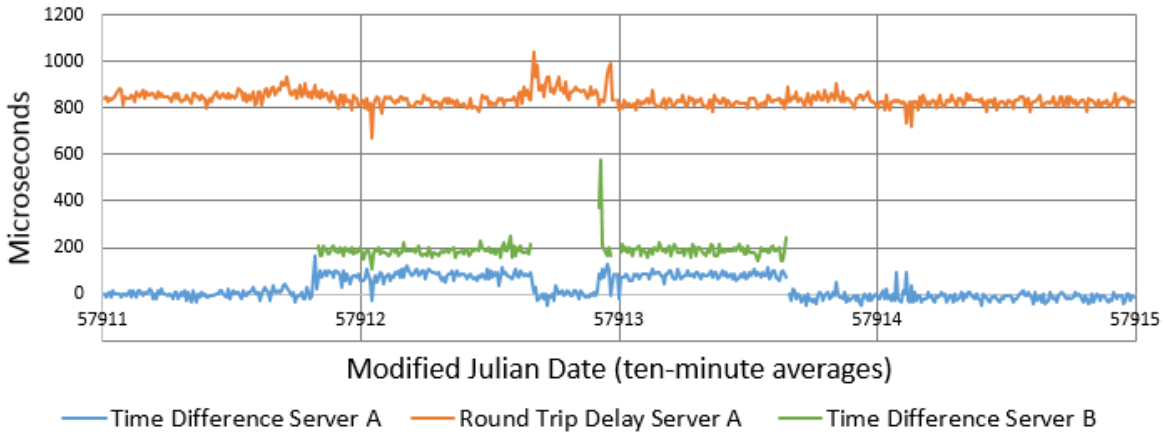


Figure 1. Two NTP servers compared to UTC(NIST) on a local network. The presence of Server B appeared to affect the time difference of Server A, but not the round-trip delay.

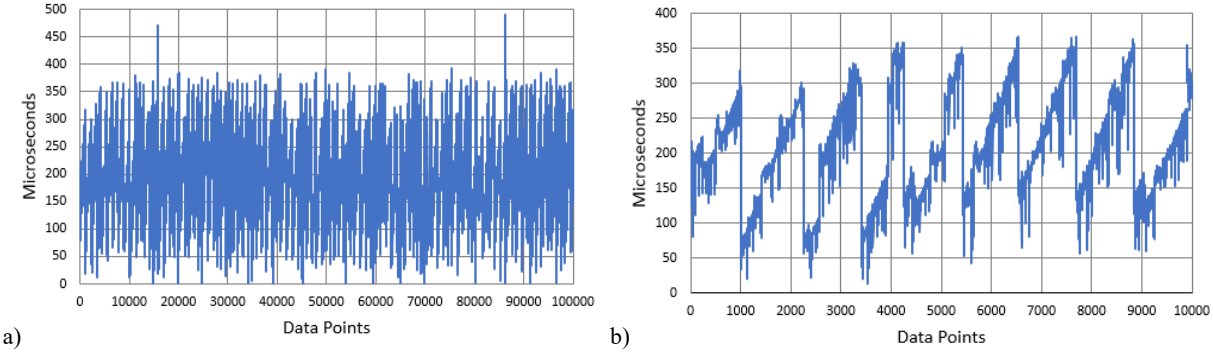
We saw the effect in the data from multiple servers by the same manufacturer. We began to investigate Server B and how it could be the source of the problem. However, we found changing it to *any* brand of server caused the servers in question to time step, even adding ones with the same make and model. After adding and removing servers, we noticed that the order of the list was significant; adding or removing a server from the list *before* the ones in question (earlier on the list) caused them to jump, but changes in the list *after* did not have an effect. Also, adding the same server to the list several times caused the problem and each successive server had a different time offset. So, it appeared that checking the servers slightly earlier or later in the 10 s segment produced different results, but if no changes to the list were made, the results remained consistent for long periods.

The root cause of this behavior was not obvious until client software was developed at NIST that made rapid packet requests of a single NTP server. The software makes user datagram protocol (UDP) requests from an NTP server as quickly as possible until it either reaches 100,000 requests or times out due to network or server limitations. The UDP port is only opened and closed once and all the data points are stored in an array and written to disk. Figure 2 shows a graph of the entire dataset and a subset of the first 10,000 points captured by the software. It shows that the NTP timing output has a "sawtooth" shape that repeats about every 1000 points. The software pushes the server to its capacity, which is specified by the manufacturer to be 1000 requests per second, and the number of points in each sawtooth varies around this number. By measuring the time interval of the complete data run, it appeared that the errors repeated every second. Thus, when the measurement system described previously checked the list of servers in the same order every 10 s, the requests to each server occurred at about the same place in the 10s cycle, hiding the true range of possible outputs from the servers in question. If this were not true, we would have seen a larger range in the data from Server A in Figure 1. Typically, an NTP client checks a server at intervals that are longer than 10 s and, on a public network, requests might reach the server at different points during the second, so the results would be spread out across the range of the sawtooth.

The range of the data shown in Figure 2 is ~400 µs, and the average time difference in our test is 211.3 µs, which is still less than the 1 ms timing specification of the NTP server and would not be noticeable on a WAN. On a LAN, however, it was apparent that the uncertainty of the transmitted packets was dependent upon when the request was made. After sharing this

Novick, Andrew; Lombardi, Michael; Franzen, Kevin; Clark, John.
"Improving packet synchronization in an NTP server."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) ,
Reston, VA, United States. January 29, 2018 - February 1, 2018.

information with the manufacturer, they performed some independent tests to try to replicate the results to determine the cause of the issue. We also pointed out the occasional outliers and questioned what their expected time offset was, because as was the case with Server A in Figure 1, it was very close to zero at times but that appeared to be coincidental.



a) b)

Figure 2. a) Time difference data from an NTP server synchronized to UTC(NIST), measured by a client system on the same subnet making rapid packet requests, and compared to UTC(NIST). b) A magnification of the first 10,000 points of 2a), revealing repetitive time steps resulting in a sawtooth shape.

3. MANUFACTURER'S SETUP

A logic analyzer was used at the manufacturer's site to test points on two identical devices, one configured as an NTP server with a Global Positioning System (GPS) input as the timing reference and one configured as an NTP client. The logic analyzer was connected to the Reduced Media Independent Interface (RMII) between the central processing unit (CPU) and the physical layer chip (PHY). This enabled it to capture the time when the NTP packet was sent/received on the Ethernet interface. On the server and the client four signals were monitored: *RX time captured* (pulse on a CPU pin when an NTP packet is received), *RMII TX* (a pin that goes high when data are transmitted), *RMII RX* (a pin goes high when data are received), *TX time inserted* (pulse on a CPU pin when placing the time stamp into the NTP packet). One channel of the logic analyzer utilized a 1 pps signal from a GPS receiver so that it had the same reference as the server. This setup allowed the analysis of the NTP request/response with 1 µs uncertainty on the logic analyzer. Packet analyzer software was used to save the packet data of each time stamp. With these measurements, they found that there was a rounding/truncation error in the code generating the time stamps.

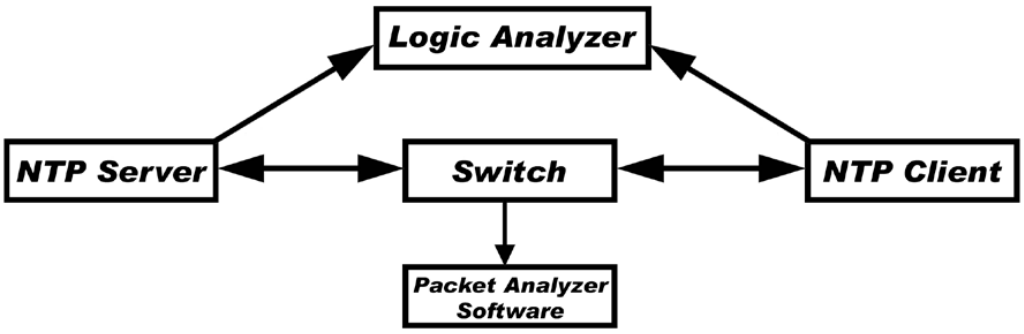


Figure 3. Block diagram of manufacturer's measurement setup.

The server code had been ported from another platform where the CPU did not have the processing power to use double floating-point math or 64-bit integers, it was limited to using 32-bit integer math. This limited the resolution of the time stamp

Novick, Andrew; Lombardi, Michael; Franzen, Kevin; Clark, John.
"Improving packet synchronization in an NTP server."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) ,
Reston, VA, United States. January 29, 2018 - February 1, 2018.

when converting from NTP fractional seconds to microseconds. The NTP fractional seconds are represented as a 32-bit binary number (with a maximum value of 4,294,967,295), but when this was divided into microseconds, only the integer value of 4,294 microseconds was used. This method discarded relevant values up to 0.967295 µs out of 4,294 counts per second, resulting in an accumulating error up to 225 microseconds during each one second period. This error is reflected in the sawtooth pattern shown in Figure 2. The current platform code was updated to utilize 64-bit integers when converting to microseconds which included all of the relevant values, eliminating the rounding error.

Looking at the occasional outliers in the NIST data, such as the outliers from Server A in the ten-minute averages on modified Julian date (MJD) 57914 in Figure 1, the manufacturer worked to minimize these occurrences. They revised thread priorities to improve the consistency and reduce the uncertainty of the NTP time stamps, raised the priority of the NTP thread and, where possible, lowered the priority of other threads so that they would be below NTP in the hierarchy. They developed a firmware update and sent it to NIST for installation and testing.

4. MEASUREMENT RESULTS

Once the new firmware was installed at NIST, the rapid packet testing was repeated and the results are shown in Figure 4. The changes removed the sawtooth from the time stamps contained in the packets and thus their uncertainty is no longer dependent on the instant when the request is made. This was confirmed by adding and subtracting servers from the list of IP addresses on the client measurement system, which had no noticeable effect on the measured time differences. The overall range from the rapid packet requests was reduced to 111.0 µs and the time difference average from 10,000 points was reduced to 34.7 µs, a significant improvement compared to Figure 2 in Section 2.

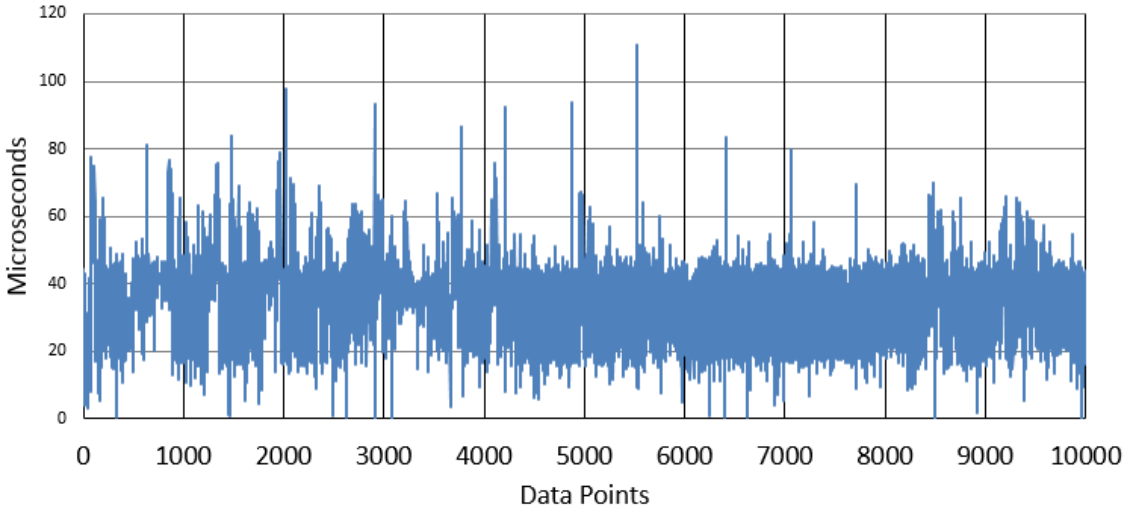


Figure 4. Time differences from an NTP server compared to UTC(NIST) obtained from rapid packet requests for ~12 s after a firmware upgrade.

The manufacturer believed that they made improvements to the packet synchronization beyond the correction of the rounding error, and they were interested in knowing if the outlying points we saw in the ten-minute averages from the client measurement system had been alleviated as well. We did not have an idea of what the uncertainty of the synchronization was before the firmware change, because it could theoretically be anywhere within the range of the sawtooth. The manufacturer created a version of the firmware that repaired the rounding error, but left out all the other improvements to the timing they had added. Figure 5 shows a comparison of the measurements using the firmware with only the rounding error fix and the final version of the firmware. The graph shows an improvement in the average time difference, from 118.0 µs to 29.4 µs for this data set. Also, the range was reduced by more than half, from 32 µs to 14 µs, the data are less noisy, and there are fewer outlying points.

Novick, Andrew; Lombardi, Michael; Franzen, Kevin; Clark, John.
"Improving packet synchronization in an NTP server."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) ,
Reston, VA, United States. January 29, 2018 - February 1, 2018.

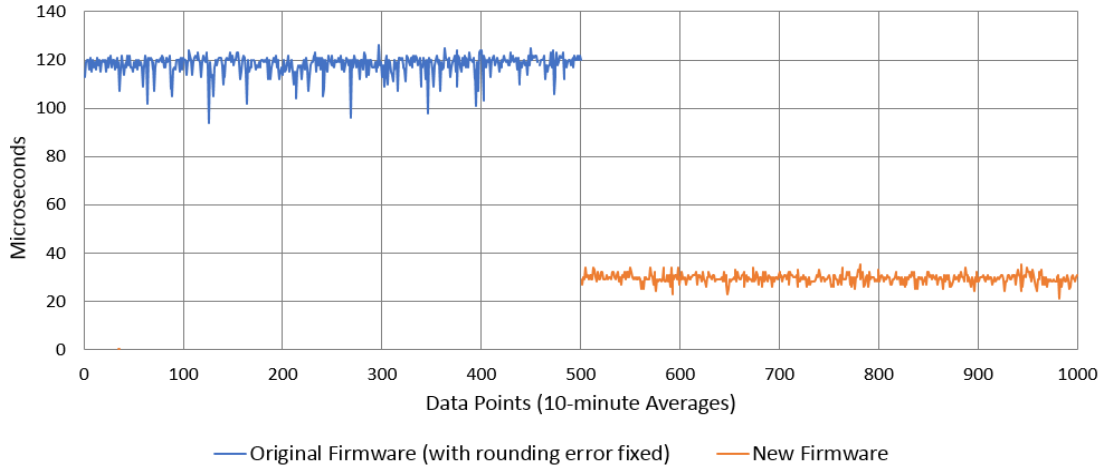


Figure 5. A comparison of the NTP server time differences between the original firmware with only the rounding error fixed, and the firmware after several improvements.

5. SUMMARY

Recent work at NIST has led to an investigation of distributing UTC(NIST) across local networks using NTP packet synchronization. We have measured the timing limitations and uncertainties caused by elements on the network, including the variation in the outputs of different servers residing on the same network. We found that server clock uncertainty can be a significant contributor to the NTP measurement uncertainty, especially on a LAN. By comparing the packet synchronization of a server referenced to UTC(NIST) to a client also referenced to UTC(NIST), we discovered periodic time steps in the server's output. When this was reported to the manufacturer, they resolved the problem and developed a firmware upgrade that greatly reduced the uncertainty of the packet synchronization.

6. REFERENCES

[1] M. Lombardi, J. Levine, J. Lopez, F. Jimenez, J. Bernard, M. Gertsvolf, et al., "International Comparisons of Network Time Protocol Servers," *Proceedings of the 2014 Precise Time and Time Interval Systems and Applications Meeting*, 1-4 December, 2014, Boston, Massachusetts, 57-66.

[2] S. Sommars, 2017, "Challenges in Time Transfer Using the Network Time Protocol (NTP)," *Proceedings of the 2017 Precise Time and Time Interval Systems and Applications Meeting*, 30 January–2 February, 2017, Monterey, California, 271-290.

[3] K. Vijayalayan and D. Veitch, 2016, "Rot at the roots? Examining public timing infrastructure," *Proceedings of the 35th Annual IEEE International Conference on Computer Communications,* 10-14 April, 2016, San Francisco, California, 1-9.

[4] A. Novick and M. Lombardi, 2015, "Practical Limitations of NTP Time Transfer," *Proceedings of the 2015 Joint Conference of the IEEE International Frequency Control Symposium and the European Frequency and Time Forum*, 12-16 April 2015, Denver, Colorado, 570-574.

[5] A. Novick and M. Lombardi, 2017, "A comparison of NTP servers connected to the same reference clock and the same network," *Proceedings of the 2017 Precise Time and Time Interval Systems and Applications Meeting*, 30 January–2 February, 2017, Monterey, California, 264-270.

Novick, Andrew; Lombardi, Michael; Franzen, Kevin; Clark, John.
"Improving packet synchronization in an NTP server."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) ,
Reston, VA, United States. January 29, 2018 - February 1, 2018.

Implementation of SDR TWSTFT in UTC Computation

Zhiheng JIANG[1], Felicitas ARIAS[1,13], Victor ZHANG[2], Yi-Jiun HUANG[3], Joseph ACHKAR[4], Dirk PIESTER[5], Shinn-Yan LIN[3], Wenjun WU[6], Andrey NAUMOV[7], Sung-hoon YANG[8], Jerzy NAWROCKI[9], Ilaria SESIA[10], Christian SCHLUNEGGER[11], Kun LIANG[12], Miho Fujieda[14]

1 BIPM: Bureau International des Poids et Mesures, Sevres, France, zjiang@bipm.org
2 NIST: National Institute of Standards and Technology, 325 Broadway, Boulder, CO80305, USA
3 TL: National Standard Time and Frequency Laboratory, Telecommunication Laboratories, Chunghwa Telecom, Taiwan
4 OP: LNE-SYRTE, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, Paris, France
5 PTB: Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany
6 NTSC: National Time Service Center, Lintong, Xian, China
7 SU: Main Metrological Center for State Service of Time and Frequency, FGUP VNIIFTRI, Russia
8 KRISS Korea Research Institute of Standards and Science
9 AOS: AOS: Astrogeodynamic Observatory of Space Research Center, Borowiec, Poland
10 IT/INRIM: Istituto Nazionale di Ricerca Metrologica, 10135 Torino, Italy
11 CH/METAS: Federal Institute of Metrology METAS, Lindenweg 50, 3003 Bern-Wabern, Switzerland
12 NIM: National institute of Metrology, 100029 Beijing, China
13 SYRTE, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, Paris; Retired from BIPM
14 NICT: National Institute of Information and Communications

BIOGRAPHIES

Zhiheng JIANG obtained his Ph.D. at the Paris Observatory. He has worked at the Research Institute of Surveying and Mapping, Beijing China; the French Geographical Institute, France and since 1998 at the International Bureau of Weights and Measures (BIPM) as a principal physicist. His main interest is in timing metrology and gravimetry, as well as being deeply involved in the generation of the international time scale, Coordinated Universal Time (UTC). Email: zjiang@bipm.org

Victor Shufang ZHANG has a M.S. degree from the Electrical and Computer Engineering Department of the University of Colorado. He has been with the Time and Frequency Division of the National Institute of Standards and Technology (NIST) since 1989, working on various GPS and TWSTFT time and frequency transfer projects and time transfer link calibrations. He served as a co-chair of the Civil GPS Service Interface Committee (CGSIC) timing subcommittee from 2002 to 2015. Starting in September 2015, he serves as chair of the CCTF Working Group on Two-Way Satellite Time and Frequency Transfer. Email: victor.zhang@nist.gov

Yi–Jiun HUANG received the Ph.D. degree in communication engineering from National Taiwan University, Chinese Taipei, in 2017. In 2009, he joined the Telecommunication Laboratories, Chunghwa Telecom Co., Ltd. as a researcher. His research interest includes satellite-based time and frequency transfer and synchronization techniques. Email: dongua@cht.com.tw

Joseph ACHKAR received the Accreditation to supervise research (HDR) in microwave metrology and time metrology from Université Paris 6 and the PhD degree in the telecommunications from Université Paris 7, in 2004 and 1991, respectively. He has been with the French national metrology institute, heading the RF and microwave standards group of BNM-LCIE (1991 – 2001) and coordinating the Time metrology group of LNE-SYRTE (2002-2017). His main research interest is focused on TWSTFT. Since 2005, he is head of delegation for France at the WP-7A meetings of ITU (he was appointed French coordinator on UTC at RA in 2012 and WRC in 2015). He acts as a technical assessor on behalf of the French (COFRAC) and Luxembourg (OLAS) accreditation bodies, since 2001 and 2008, respectively. Email: joseph.achkar@obspm.fr

Dirk PIESTER received his diploma degree in physics and Dr.-Ing. degree from the Technische Universität Braunschweig. Since 2002, he has been with the Time Dissemination Group of the Physikalisch-Technische Bundesanstalt (PTB) where he is in charge of PTB's time services. His main research interests are improvements of time and frequency transfer technologies and calibration techniques via telecommunication and navigation satellites as well as optical fibers. In this framework two fellowships (in 2007 and 2010) were granted by the Japan Society for the Promotion of Science (JSPS). Dr. Piester has served as the chair of the CCTF Working Group on Two-Way Satellite Time and Frequency Transfer from September 2009 to September 2015. He has authored or co-authored more than 90 papers in refereed journals and conference proceedings. Email: dirk.piester@ptb.de

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

Shinn-Yan LIN (**Calvin**) was born in Chinese Taipei, in 1967. He received a B.S. in physics and a M.S. degree in astronomy physics from the National Central University, Chung-Li, in 1990 and 1994. He then entered the National Standard Time and Frequency Laboratory, Telecommunication Laboratories, Chunghwa Telecom Co. Ltd, Chinese Taipei, as a Researcher. He works on the national standard time and frequency project and is in charge of the maintenance and improvement of the national standard time scale of Chinese Taipei. His current research interests include timescale algorithms, GNSS and satellite time and frequency transfer, local time dissemination systems, and precise time and frequency measurements. Email: sylin@cht.com.tw

Wenjun WU received his PhD degree in 2012 from the Graduate University of Chinese Academy of Sciences and his work mainly focused on TWSTFT. He worked as a visiting scientist for the UTC time link calibration at BIPM from June 2014 to May 2015. Now he is an associate researcher in National Time Service Center, Chinese Academy of Sciences (NTSC), and his main research fields are TWSTFT and GNSS time transfer. Email: wuwj@ntsc.ac.cn

Ilaria SESIA received an M.S. degree in Electrical Engineering from Institut National Polytechnique de Toulouse, an M.S. degree in Communication Engineering and a PhD degree in Metrology from Politecnico di Torino. She is a Researcher at the Italian National Metrology Institute (INRIM), where she is involved in metrological timekeeping activities, mainly working on the characterization of atomic clocks and time scales in satellite applications. She is in charge of INRIM time and frequency transfer via geo-stationary satellite. Since 2004, she has been deeply involved in the design and development of the European satellite navigation system Galileo. Email: i.sesia@inrim.it

ABSTRACT

Two-Way Satellite Time and Frequency Transfer (TWSTFT or TW for short) is one of the primary techniques for the realization of Coordinated Universal Time (UTC), which is computed by the International Bureau of Weights and Measures (BIPM). TWSTFT is carried out continuously by about 20 timing laboratories around the world. One limiting factor of the TWSTFT performance is the daily pattern (diurnals) in the TWSTFT data. Their peak-to-peak variations were observed as up to 2 ns in some extreme cases. They must be attributed to the equipment on ground or the satellite transponder, but no clear understanding has been achieved for some years. In 2014 and 2015, it was demonstrated inlinks between Asian stations that the use of Software-Defined Radio (SDR) receivers for TWSTFT could considerably reduce the diurnals and also the TWSTFT measurement noise.

In 2016, BIPM and the Consultative Committee for Time and Frequency (CCTF) working group (WG) on TWSTFT launched a pilot study on the application of SDR receivers in the Asia to Asia, Asia to Europe, Europe to Europe and Europe to USA TWSTFT networks.

The very first results of the pilot study have been reported to the PTTI 2017. The results show, 1) for continental (short) links: SDR TWSTFT demonstrates a significant gain in reducing the diurnals by a factor of two to three; 2) for inter-continental (very-long) links: SDR TWSTFT displays a small gain of measurement noise at short averaging times; 3) SDR receivers show superior or at least similar performance compared with SATRE† measurements for all links.

The CCTF WG on TWSTFT prepared a recommendation "On Improving the uncertainty of Two-Way Satellite Time and Frequency Transfer (TWSTFT) in UTC Generation" for the 21st CCTF meeting. One of the recommended items is to introduce the use of SDR TWSTFT in UTC generation. The recommendation was approved by CCTF during the meeting in June 2017.

In the current setups, a SDR receiver and the collocated SATRE modem receive the signals transmitted by remote SATRE modems, and the two devices independently determine the arrival time of the received signal. Thus, a few setup changes are necessary to implement SDR receivers into operational TWSTFT ground stations. On the other hand, the data computation, e.g. for calibrated time transfer, needs some caution in the data processing and provision, which are not trivial. Thus, an Ad hoc Group has been established to work out a procedure for the use of the SDR TWSTFT in UTC computation.

This paper reports on the progress of the work for using SDR TWSTFT in UTC computation. Section 1 introduces SDR TWSTFT and the pilot study. Section 2 presents the considerations for using SDR TWSTFT in UTC generation. Section 3 and 4 show the methods for analysing SDR TWSTFT and the analysis results. Sections 5 and 6 discusses the further improvement of SDR TWSTFT and summarizes the work towards using SDR TWSTFT in UTC generation.

1. INTRODUCTION

Two-Way Satellite Time and Frequency Transfer (TWSTFT or TW for short) is a primary technique for the generation of Coordinated Universal Time (UTC) [1,2]. Its dominant source of statistical uncertainty is the daily variation (diurnals) observed in the link data, whose peak-to-peak amplitude can be up to 2 ns in extreme cases.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

SDR TWSTFT stands for the application of Software-Defined Radio receivers in TWSTFT. A SDR receiver (SDR for short) works in parallel with the Satellite Time and Ranging Equipment (SATRE modem or SATRE for short) which has been used by the international timing community since 2003 in the UTC generation. In fact, the SDR receives the down converted signal (Rx) and then determines the arrival time of the signal independent of the SATRE [3], cf. the Figure 1.1. With the application of SDR in TWSTFT, a reduction of the diurnals and thus the measurement noise in Asian TWSTFT links was observed [3,4]. Because of the encouraging results and after a first validation by the International Bureau of Weights and Measures (BIPM) [4], the BIPM and the Consultative Committee for Time and Frequency (CCTF) working group (WG) on TWSTFT launched a pilot study during the TWSTFT participating stations meeting at the Precise Time and Time Interval (PTTI) conference in Monterey, California in February 2016. The goal of the pilot study has been to investigate the impact of using SDR TWSTFT in Asia to Asia, Asia to Europe, Europe to Europe and Europe to USA links with different satellites, c.f. the Figure 1.2.

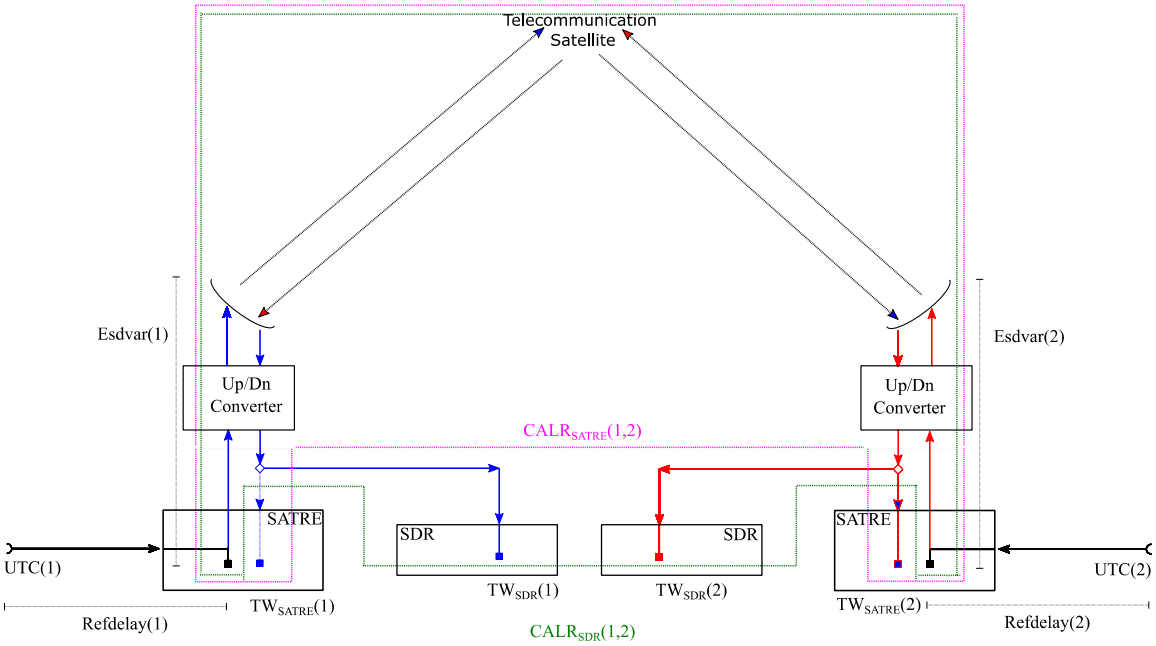


Figure 1.1 Illustration of the main paths of a TWSTFT link based on SATRE modems and on SDR receivers for comparison of two remote clocks

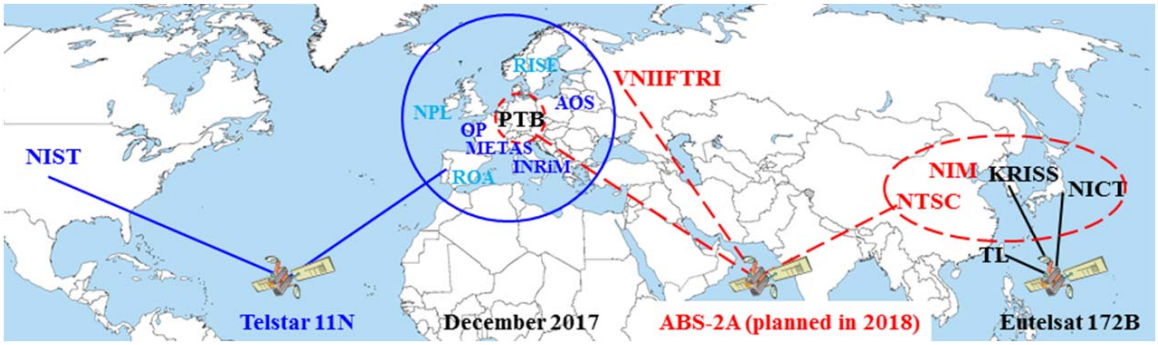


Figure 1.2 Participating stations in the SDR pilot project (status of December 2017). The links in blue are for the Europe-Europe and Europe-USA SDR TWSTFT. The links in red are for the Asia-Europe SDR TWSTFT. The links in black are for the Asia-Asia SDR TWSTFT. Until June 2017, the AM22 was used in the place of ABS-2A.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

So far, 15 timing laboratories have installed SDR receivers and participate in daily SDR TWSTFT. The participants include: TL, NICT, KRISS (KRIS), NTSC and NIM in Asia; PTB, OP, VNIIFTRI (SU), INRIM (IT), METAS (CH), AOS (PL), RISE (ex SP, Sweden), ROA (ES) and NPL (UK) in Europe; and NIST in USA. Three satellite links in the Ku band have been used in the pilot study, they were routed through the satellites Eutelsat 172B for the Asia to Asia links, Express AM22 for the Asia to Europe links until June 2017, and Telstar 11N for the Europe to Europe and the Europe to USA links. At the time of this writing, the successor of AM22 satellite has been selected as ABS-2A and the Asia to Europe TWSTFT links will soon be reinstalled.

Continental (in Asia and in Europe), and also transcontinental (Asia-Europe, Europe-USA) SDR TWSTFT time links have been established. The SDR TWSTFT pilot study has been a global experiment involving leading laboratories. During the 25th annual meeting of the CCTF WG on TWSTFT held at NTSC in May 2017, the participating laboratories reported their results and analysis of the SDR pilot study [5]. The BIPM presented its validation [6].

The major conclusions were:

- for continental links: SDR TWSTFT demonstrates significant gain in reducing the diurnals by a factor of two to three;
- for inter-continental (very-long) links: SDR TWSTFT displays little gain of measurement noise at short averaging times;
- SDR receivers show superior or at least similar performance compared with SATRE measurements for all links.

Based on these findings, the WG on TWSTFT prepared the recommendation "On Improving the uncertainty of Two-Way Satellite Time and Frequency Transfer (TWSTFT) for UTC Generation" [13] for the 21st CCTF meeting held on 9-10 June 2017 at BIPM. At the meeting, the recommendation was submitted to the CCTF and the latter approved the recommendation. This means, that the process of implementation of SDR TWSTFT into UTC generation can start.

Aiming at a routine computation of using SDR TWSTFT for UTC, there were still some open points to address. An Ad hoc Group composed of 5 experts from BIPM, NIST, OP, PTB and TL was setup to resolve these issues, comprising of data collection, treating discontinuities, calibration of SDR TWSTFT links, data format and computation programs, etc. Its attention was focussed on the practical issues towards its use in the UTC computation. The BIPM should perform a final evaluation using at least six months' continuous data. One of the tasks of the Ad hoc Group has been to prepare a report on implementing the recommendation and request approval of the implementation of the SDR TWSTFT used in the UTC generation. In this paper, the main results of the report are presented to the scientific community.

2. BASIC CONSIDERATIONS

2.1 Calibration of SDR TWSTFT links

The use of SDR TWSTFT links in UTC generation requires that the signal delays along the full propagation path have been determined. According to the TWSTFT calibration guidelines [2], a SDR link calibration can be made using a '2-step procedure' as described below. The calibration results are going to be computed and issued by BIPM coordinated with the SDR TWSTFT participants.

The two-step procedure:

1. For UTC links:
   a) if the SATRE link is calibrated, the SDR link should be aligned to it;
   b) if the SATRE link is not calibrated, both links – the SATRE and the corresponding SDR link – should be calibrated together by using a TWSTFT mobile station or GPS calibrator with the 'link' method [2];
2. The non-UTC link delay should be calibrated by the triangle closure calibration (TCC) method [2];

A Calibration Identifier (CI) number is issued to each SDR link calibration. The TWSTFT calibration guideline is to be accordingly updated.

2.2 Conventional uncertainty u_A and u_B

For the UTC SATRE links, the *conventional* values of u_A and u_B are 0.5 ns and 1.0 ns (using TWSTFT mobile station calibration) or 1.5 ns (with GPS Calibrator) respectively. These values are assigned to time differences UTC-UTC(k), reported in the BIPM Circular T (Section 1) at a five-day grid, and are derived from a detailed report on the used link data in Section 5 of Circular T. The conventional uncertainty as assigned by BIPM is usually higher than the individually estimated one. For example, the reported u_B of the USNO-PTB SATRE link calibration is 0.6 ns while the conventional value is 1.0 ns [1].

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

The corresponding values of u_A and u_B for SDR links are 0.2 ns and 1.0 to 1.5 ns, respectively. In fact, as reported in [3,5,6] and in this paper, for short baselines, such as OP-PTB, a gain factor of two or three is obtained in the stabilities (SDR vs. SATRE). For long baselines, such as NIST-PTB with much smaller diurnal comparing to most of other links, e.g. the link of OP-PTB. The stability of the SATRE link is already below to 0.2 ns, see [6] and the discussion in the section below. As for u_B, the same conventional value is valid for both the SATRE and the SDR links because, as mentioned above, the conventional u_B of the SATRE link is often higher than the measured one. Hence, the combined u_B of the SDR link is usually within the conventional u_B of the SATRE link.

The conventional value of the u_B for the SDR TCC calibration is 2 ns, the same as that when using SATRE data [2,17].The u_B of SDR links may be improved if a direct SDR link calibration is performed. This study is ongoing.

2.3 Reporting the SDR TWSTFT data

1) We focus on the data files in the format of the International Telecommunication Union Recommendation ITU-R TF.1153-4 [7] (ITU format or ITU for short) used in the UTC computation. It is recommended that the 1-s data files should be archived in local computers but not submitted to BIPM;
2) The SDR data in ITU format should be submitted to the BIPM ftp site in the dedicated SDR folder, in the same way as the SATRE TWSTFT data;
3) The conventional ITU data format and file name [7] for the SATRE modem should be used for the SDR data. The differences between a SATRE data file and a SDR data file are as the followings:
   - In the header of a SDR file, there is an identification comment line, such as:
     `* MODEM SATRE 076, Software-Defined Radio Receiver`
   - The identification number for the SDR receivers should equal the SATRE earth station number plus 50, usually to be between 51 and 59. For example, for the SDR link of NIST-OP, we have the SDR Local Station ID and Remote Station ID as 'NIST**51** OP**51'** in comparison to the SATRE Local Station ID and Remote Station ID 'NIST**01** OP**01'.**
4) For reporting the SDR data in ITU format for the UTC computation, a 300 s-interval is to be used. One of the advantages of this interval is that the SDR TWSTFT data can be compared exactly to the measurement of other time and frequency transfer data on the same epoch, such as the BIPM GPS PPP (GPS Precise Point Positioning) solution.

2.4 Further improvement on SDR TWSTFT for UTC

Without touching the hardware, we can still improve the stability of a SDR link by the following methods:

1) Combination of SDR TWSTFT and GPS PPP. The same kind of combination of SATRE TWSTFT and GPS PPP has been used in UTC time transfer since 2009 [14];
2) Complete or partial network time transfer to fully use the redundant data [8-12,15]. Previous studies [10,11] suggest a simplified version, the so-called indirect link.

The related programs of the two methods are almost available in the BIPM UTC/TAI software package Tsoft, c.f., Section 5.

2.5 Operational Procedures

The procedures listed below were defined in order to make the automatic computation of a SDR link becoming equal to that of the SATRE link in the standard BIPM Circular T monthly computation, as well as in the complete computations such as the link comparison and the publication in the BIPM web site.

1) Elimination of phase steps in the SDR measurements when a SDR receiver is restarted has to happen. The new release of the SDR software V2018.1 [25] guaranties that the measurements continue without steps when restarting the device;
2) The value and the definition of the REFDELAY (reference delay) for SDR are identical to that of the SATRE;
3) The SDR ITU format data interval is to be a fixed value, at present, 300 s. This is realised by the software obs2ITU which converts the SDR raw measurement data to the standard ITU data format;
4) CALR (link calibration) and ESDVAR (earth station delay variation) values are defined as Figure 1.1 shows. CALR and/or ESDVAR parameters may be adjusted to align SDR and SATRE links.

2.6 Schedule of the implementation of SDR TWSTFT in UTC

1) The SDR TWSTFT participants were requested to update their procedures as listed in Section 2.5 and to make the equipment operational for routine measurements before July 2017;

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

2) During July-December 2017, measurements were carried out among the SDR TWSTFT participants and the test computations/calibrations were made at BIPM;
3) BIPM made a final data analysis using the 6-month data of 1707-1712 (July 2017 to December 2017) and reports the results to the SDR ad hoc Group, and informs the WG on TWSTFT. The Ad hoc Group should prepare the final validation report to the WG and to the BIPM;
4) From October 2017 onwards, BIPM has used the SDR data as the *backup* UTC links in the monthly Circular T computation. The BIPM publishes monthly on the BIPM web the results of the SDR link and its comparison to other techniques;
5) SDR TWSTFT is going to be introduced for UTC computation in BIPM Circular T on an earliest possible date in 2018. The decision will be made jointly by the BIPM and the WG on TWSTFT, represented by the Ad hoc group, in a link by link approval;
6) In June 2018, a final report of the SDR TWSTFT Pilot Study will be presented to the 26th TWSTFT WG meeting in Poland (2018).

3. THE METHOD AND THE EARLIER STUDY

The SDR TWSTFT measurement data were collected from TL, KRISS, NICT since October 2014, and from NTSC, PTB, OP, NIST, AOS, CH(METAS), IT(INRIM), SP(RISE) and SU(VNIIFTRI) since July 2016. The SDR links discussed in Section 4 were calibrated with an alignment to the corresponding SATRE links.

In the following sections, we first make a quick review of the precedent studies and results (Section 3). We then focus on the results of the latest 6-month analysis (Section 4). In Section 5, we study and validate the methods to further improve the quality of SDR. Section 6 is a summary.

In this section, we give an outline of the analysis method and the earlier study results. Details can be found in the [5,6,17,22].The study is based on the analysis of the double clock difference (DCD) over a baseline between (a) the SATRE and the SDR links, and (b) the SATRE or SDR and GPS PPP/IPPP (Integer Ambiguity GPS PPP solution [21]) links. The DCD is very helpful for the uncertainty analysis.

Methods of the BIPM validation:

- We analyse the $\boldsymbol{\sigma}$ (standard deviation) and the Time Deviation (TDev) $\boldsymbol{\sigma}_X$ of DCD;
- The data used are mainly: SATRE TWSTFT, SDR TWSTFT, GPS PPP and GPS IPPP;
- The 4 indicators of the assessment of the quality and the level of the improvement are the following:
  1. σ_X for evaluation of the instabilities at different averaging times (τ)
  2. σ (DCD) for revealing the agreement with GPS PPP/IPPP: the smaller the better
  3. σ of the triangle closures for indication of the uncertainty and the noise level of the links: the smaller the better
  4. σ of discrepancies to GPS PPP or to BIPM Circular T for the long-term stability: the smaller the better
- Gain factors: $\boldsymbol{\sigma}_{SATRE}/\boldsymbol{\sigma}_{SDR}$, $\boldsymbol{\sigma}_{X\text{-}SATRE}/\boldsymbol{\sigma}_{X\text{-}SDR}$: the larger the better

In the future, we will have TWOTFT (Two-Way Optical-fibre Time and Frequency Transfer) as a new tool for the analysis, whose stability and accuracy are both in order of 100 ps, thanks to the links of TWSTFT and TWOTFT established between AOS and GUM (PL) [19].

Example (1) The gain in SDR link vs. the SATRE link over the baseline OP-PTB

The left plot in Figure 3.1 depicts the TWSTFT differences of SATRE and SDR links, respectively. Obviously, the SATRE link suffers considerable diurnal variations. The right plot is the TDev of the two sets of link data which shows more clearly how the diurnal and the noise level affect the SATRE TWSTFT from 2 hour to about one day. Table 3.1 gives the gain factor in TDev, 3.7 on average.

Table 3.1 Gain factors in TDev of the SDR versus SATRE link at different averaging time (τ)

τ /h	Gain
2	3.3
8	5.2
16	3.9
mean	**3.7**

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

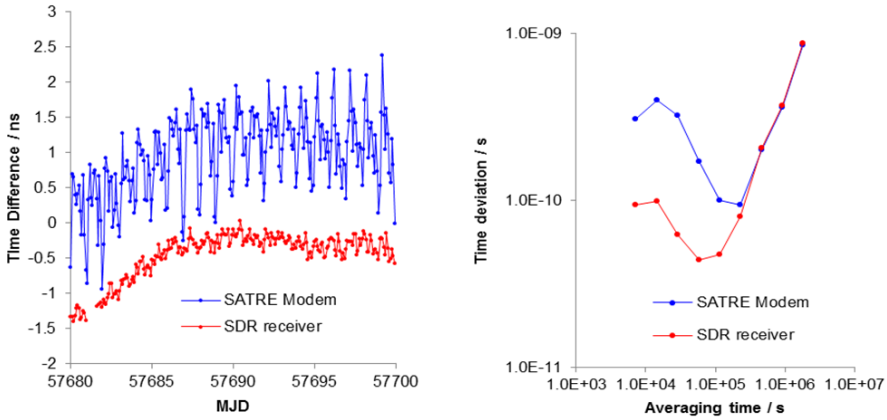


Figure 3.1 The SDR and the SATRE links and the corresponding TDev over the baseline OP-PTB

Example (2) Comparing the GPS PPP link to the TW SATRE and TW SDR links to study the gains

In the following we will compare GPS PPP links to TW SATRE and TW SDR links over the baselines TL-NICT and OP-PTB in Asia and in Europe using different satellites.

Unlike to example (1) where the TW SATRE solution was used as reference, we use the GPS PPP as the reference. Being independent from TWSTFT, the GPS PPP solution seems suitable to validate the improvement in the SDR method. Its statistical measurement uncertainty is small (u_A=0.3 ns [1]) and it is almost not affected by diurnal variations. Making a choice between TW SATRE or TW SDR link data, those which agree better with GPS PPP are considered to better represent the clock differences. Comparing TWSTFT to GPS PPP allows for also investigating the stability of SDR in view of the calibration, a key issue in the UTC time transfer. We emphasize that SDR TWSTFT and GPS PPP are independent and therefore the standard deviation and the TDev of the DCD give a conservative estimation than the truth.

Table 3.2 Gain factor in TDev in the DCD of SATRE versus SDR links against the PPP link over the baseline TL-NICT

τ /h	SATRE-PPP σ_X /ps	SDR-PPP σ_X /ps	Gain $\sigma_{X/SATRE-PPP}/\sigma_{X/SDR-PPP}$
1	140	40	3.5
2	90	45	2.0
6	125	38	3.3
12	79	48	1.6
24	68	59	1.2
Mean			**2.3**

Table 3.3 Gain factor in TDev in the DCD of SATRE versus SDR link against the PPP link over the baseline OP-PTB

τ /h	SATRE-PPP σ_X /ps	SDR-PPP σ_X /ps	Gain $\sigma_{X/SATRE-PPP}/\sigma_{X/SDR-PPP}$
2	210	160	1.3
4	340	160	2.1
8	330	110	3.0
16	170	80	2.1
Mean			**2.1**

Tables 3.2 and 3.3 contain the gain factors in TDev in the DCD of SATRE versus SDR links against the PPP link on different averaging times ranging from 2 hours to one day. The average gain is 2.3 for the baseline TL-NICT and 2.1 for OP-PTB, considering that PPP is not errorless.

Example (3) Comparison of the triangle closures given by the SATRE and SDR measurements

In theory, the triangle closure of (A-B) + (B-C) - (A-C) should be zero. A non-zero closure is hence a 'real' error, the larger the worse, the smaller the better. The gain factor is here the ratio of the standard deviation of closures. Table 3.4 gives the result of three triangles. The gain factor is 4.5 on average. It is important to point out that the deviation of the closure results constitutes

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

a real error, therefore the gain factor here indicates potentially a significant improvement of the systematic uncertainties in TWSTFT links. C.f. the Section 4.3 of TM267 [6] for detailed information on the data used.

Table 3.4 Gain factor of the SDR vs. the SATRE links given by the triangle closure analysis

Triangles	Gain factor= σ(SATRE)/σ(SDR)
OP-NIST-PTB	4.3
NICT-KRIS-TL	4.8
PTB-NTSC-SU	4.5
Mean	**4.5**

Example (4) Long-term stability of the SDR TWSTFT links

Then long-term stability of SDR links is no doubt one of the most important issues. Once a SDR link is calibrated, it should be kept stable. A comparison between the SDR TWSTFT and the GPS PPP over the baseline TL-KRIS for about 400 days was made and the σ of the DCD is 1.1 ns, which is the usual long-term variation or even better between the TWSTFT and GPS. The long-term stability of the SDR link is satisfactory for use as UTC time transfer. C.f. the Section 4.4 of [6] for detailed information on the data used.

Example (5) Test calibration of the SDR Europe-USA network

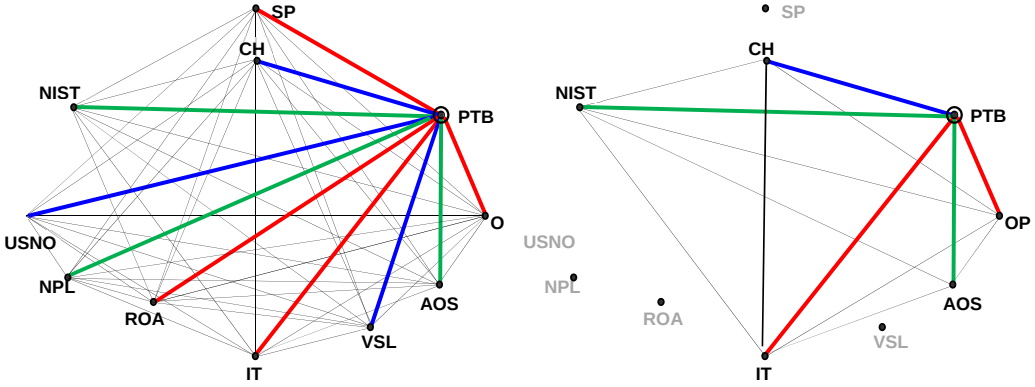


Figure 3.2 The Europe-USA TW SATRE (left) and TW SDR (right) networks

The test calibration result is given in Table 3.5 of the TM273 [23]. The data set used was the UTC 1704 (April 2017). This is just a numerical experience and the CI is assigned 999 for all the links.

Example (6) Further improve the quality of the SDR TWSTFT

Further improving the stability of the SDR is possible. Practically, we have immediately two numerical tools: the combination of SDR TWSTFT and GPS carrier phase [14] and the indirect links [8,10,11]. Figure 3.3 shows the TDev of the PTB-OP TWSTFT differences computed with the SATRE, SDR and the *indirect*-SDR via NIST. The numerical tests show that the latter is the most stable, cf. the next section for the improvement with the combination of SDR TWSTFT and GPS PPP.

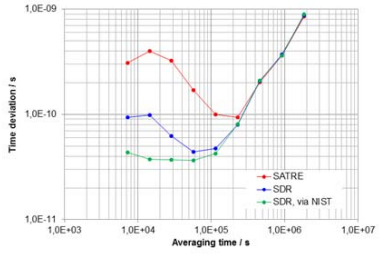


Figure 3.3 TDev of the TWSTFT link PTB-OP computed with the SATRE, SDR and the *indirect*-SDR methods

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

4. LATEST ANALYSIS USING 6-MONTH DATA OF JULY-DECEMBER 2017

Required by the CCTF WG on TWSTFT, the BIPM performed a final analysis aimed at verifying that the procedures discussed in Section 2 are adequate for using the SDR links in UTC generation. 183 days of data from the latest 6 months (July to December 2017, MJD 57935-58118) were used in the analysis. Because the Asia-Europe TW links were not operational during this period due to the AM22 satellite being out of service, only the available Europe-Europe, Europe-America and Asia-Asia links have been analysed in this report, c.f. the Figure 1.2.

In addition to the SDR TWSTFT data, the other data analysed were collected from the UTC data sets. A total of four types of the measurements are collected and carefully analysed, that include SATRE TWSTFT, SDR TWSTFT, GPS PPP and GPS IPPP. At this moment, only the OP-PTB GPS IPPP link was available.

4.1 Calibration

The SDR links analysed here were relatively calibrated by alignment to the related SATRE links using the data set 1709 (September 2017). Table 4.1.1 below lists the SDR TWSTFT calibration corrections for the UTC links in Europe and a non-UTC link in Asia. The calibration of the TL-KRIS SDR link was made by the alignment to the GPS PPP link. Here σ is the standard deviation of the alignment, that is, the fitting to the measurements that gives the scatter level of the measurements containing the noises from both the SATRE and SDR. It is not the standard deviation of the mean values (the correction) which can be computed by the σ over the root of *N*-1. The 'alignment' uncertainty is in fact very small and negligible in the 2-step calibration procedure using SDR and the SATRE links.

Note here that, the calibration corrections given in Table 4.1.1 are of the new setups and the latest values used for the UTC backup computation. Here yymm stands for year and month. The same is in the following discussion.

Table 4.1.1 The TW SDR link calibration corrections

Link	Correction /ns	N-Point	σ /ns	CI of SATRE link	Data set used /yymm
AOS-PTB	-54.449	329	0.659	449	1709
CH-PTB	-128.909	356	0.454	284	1709
IT-PTB	-188.926	376	0.519	434	1709
NIST-PTB	-113.862	245	0.108	393	1709
OP-PTB	2263.247	305	0.428	437	1709
TL-KRIS	-2441.649	4101	0.241	GPS PPP	1710

4.2 One-month data analysis

UTC is computed and published monthly. The monthly data set is therefore the base of the analysis. In this section, we investigate all the SDR links for UTC backup computation. The data were collected from the latest data sets 1712 (December 2017). The GPS IPPP links are not calibrated.

Below, we will discuss the time links, the link comparisons, that is the DCD, between SDR, SATRE, GPS PPP and GPS IPPP, as well as the related statistics. We discuss the daily and monthly stabilities and the biases between techniques. Special attention is paid to the reduction of the diurnals which may be the most important advantage of the SDR technique.

The analyses of 1-month data and 6-month data (Section 4.3) reveal the major issue of the SDR TWSTFT at this time: the discontinuity (missing data and time step) due to the changes in hardware, software and reference signals. These issues can be mitigated by improvement of the SDR TWSTFT system and carefully monitoring the SDR TWSTFT operation.

In the following, we will not show all the results but some typical examples.

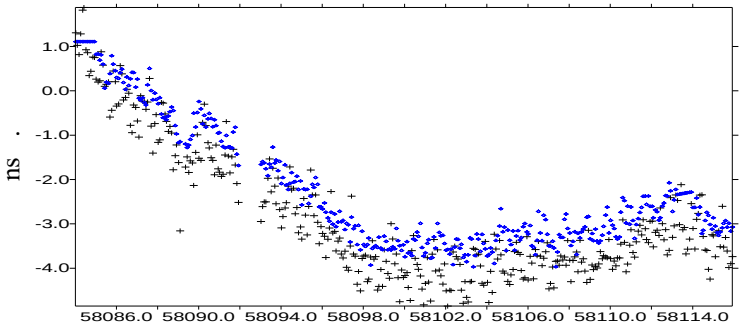
4.2.1 The OP-PTB baseline

In Figure 4.2.1x, DCD is obtained by differentiating two time link data sets. The TDev plots show the log σ_x (in seconds) versus log averaging time (in seconds). "-10" on the vertical represents thus 100 ps. The numbers in the plots represent σ_x-values in picoseconds.

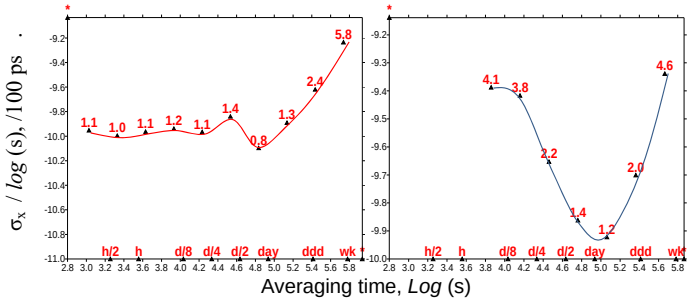
We first compare the SDR link to the SATRE link and the results are given in Figure 4.2.1a. Then we compare the TW SDR link to the GPS IPPP link. The IPPP solution is the GPS PPP solution obtained by solving the inter ambiguity. It is more precise than the PPP solution. The results are given in Figures 4.2.1b-1 to 4.2.1b-3. Because the IPPP solution is more precise than the PPP one, GPS PPP is not used for the analysis over the baseline OP-PTB when IPPP is available.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

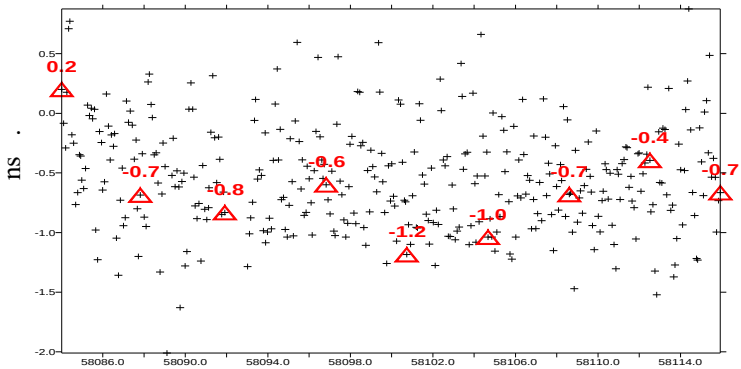
A) Comparing the SDR TWSTFT to SATRE TWSTFT



(a) SDR TWSTFT (blue) and SATRE TWSTFT (black) / MJD



(b) Time Deviations of the SDR (left) and SATRE (right) TWSTFT links



(c) DCD of SDR link and SATRE link with σ(DCD)=0.517 ns

Figure 4.2.1a Comparison between TW SDR and TW SATRE time links over the baseline OP-PTB

B) Comparing the SDR TWSTFT to GPS IPPP

The best and most accurate reference to validate the SDR technique should be the TWOTFT (Two-Way Optical-fibre Time and Frequency Transfer), which has the stability and accuracy of about 100 ps [19]. Unfortunately, such a baseline where the two techniques of SDR TWSTFT and the TWOTFT are available does not exist at this time. The next most precise technique for time transfer at present is the IPPP, i.e. the GPS PPP with integer ambiguity [21]. Both the TWOTFT and the IPPP are believed free of the disturbances seen as diurnal variations.

As well known, the phase ambiguities of GPS carrier frequencies are physically integer number but the usual PPP solves the ambiguities as real values. Associated errors may sum up as Random Walk instability. The IPPP may eliminate this bias. When required, the BIPM produces the IPPP solution using the CNES/GRGS specific satellite products and software GINS.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

Experience shows that at averaging times of 4 to 5 hours, 2×10^{-15} in frequency stability may be reached. Comparison with the TWOTFT link indicates an accuracy of frequency transfer of 1×10^{-16} may be reached in 3 to 5 days. The frequency transfer of the 10^{-16} level has been reported by comparing with two-way carrier phase [24].

We have a 10-day of IPPP solution for MJDs from 57935 to 57944 in the beginning of July. Here after we show the link and link comparison results.

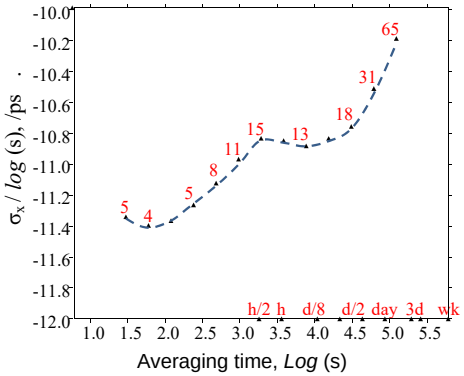


Figure 4.2.1b-1 The TDev of the IPPP time link over the baseline OP-PTB

A comparison between GPS PPP and GPS IPPP time links over the baseline OP-PTB gives that the σ(DCD)=0.088 ns. The PPP and IPPP solutions are considered to be correlated. The latter is more precise.

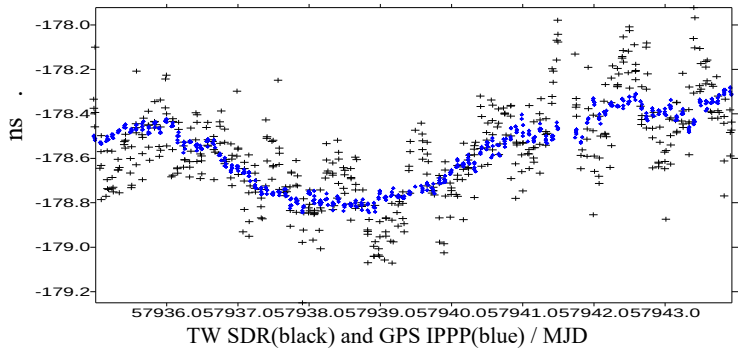


Figure 4.2.1b-2 Comparison between TW SDR and GPS IPPP time links over the baseline OP-PTB with σ(DCD)=0.158 ns

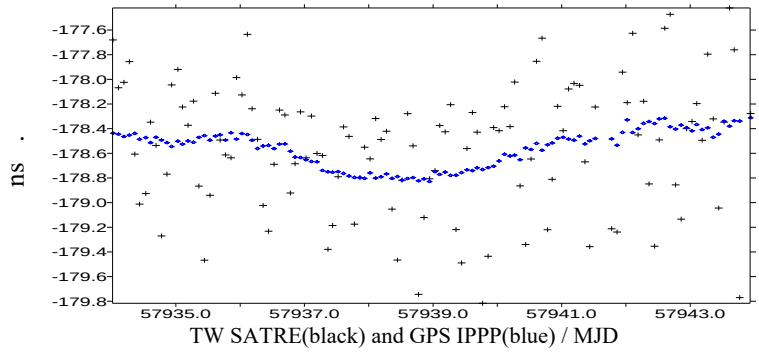


Figure 4.2.1b-3 Comparison between SATRE and IPPP time links over the baseline OP-PTB with σ(DCD)=0.516 ns

4.2.2 The TL-KRIS baseline

Similar to the discussion in the last section, we show the SDR time link and the TDev over the baseline TL-KRIS in the following figure. As can be seen, the instability of the SDR link reaches 45 ps in less than 60 minutes of averaging time.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

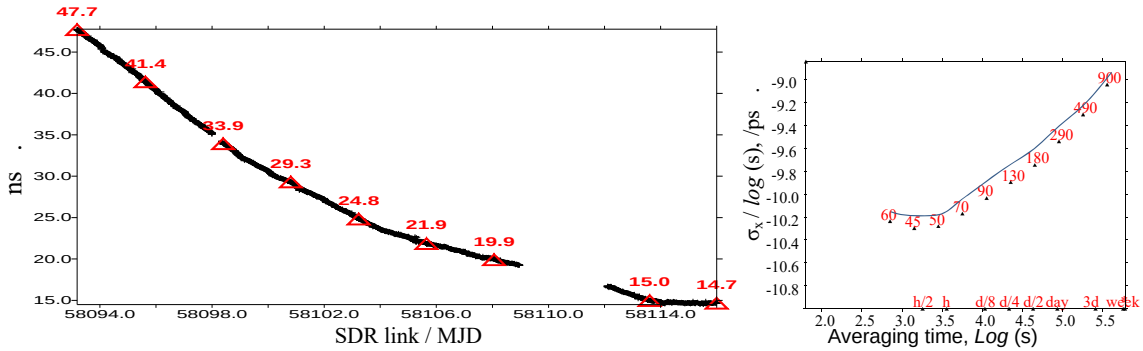


Figure 4.2.2 SDR link and the TDev over the baseline TL-KRIS

The SDR link can be further improved by the methods of the indirect linking and the combination of the SDR and GPS PPP. See details in the Section 5 below.

4.2.3 The AOS-PTB baseline

Figure 4.2.3 shows the SDR and SATRE time links and the link comparison. The σ(DCD)=0.727 ns. As can be seen obviously, the instability comes mainly from the noise in the SATRE measurements.

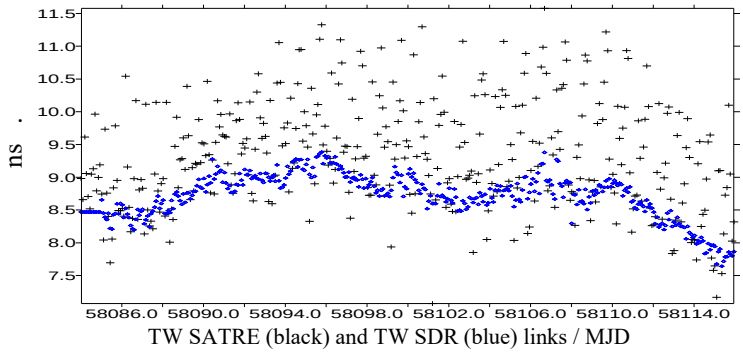
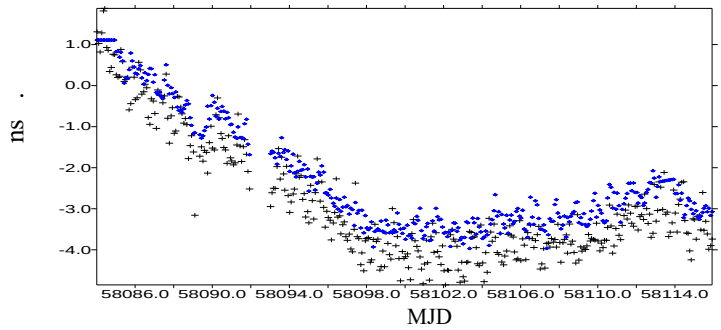


Figure 4.2.3 Comparison between SDR and SATRE time links over the baseline AOS-PTB with σ(DCD)=0.727 ns.

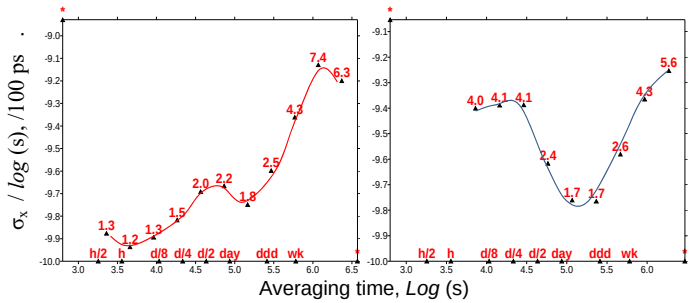
4.2.4 The IT-PTB baseline

The same as above, we show the SDR vs. the SATRE time links and the TDev over the IT-PTB baseline. As can be seen in the TDev plots the stability in the SDR is largely better than that of the SATRE at any averaging time.



(a) SATRE (black) and SDR (blue) links with the σ(DCD)=0.443 ns

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.



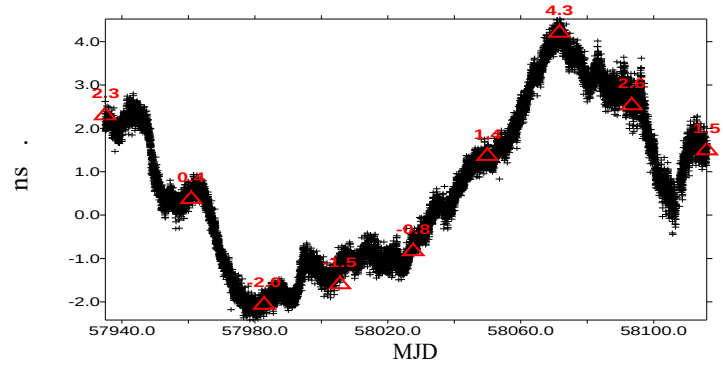
(b) TDev of SDR (left) and SATRE (right) links

Figure 4.2.4 Comparison between TW SDR and TW SATRE time links over the baseline IT-PTB

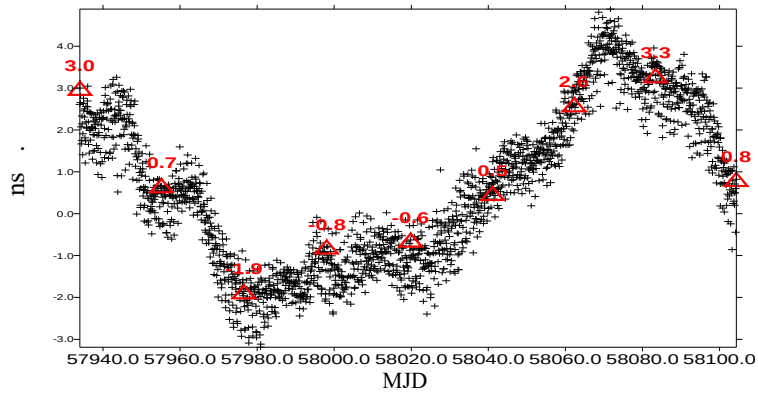
4.3 Six-months data analysis

Obviously, the long-term stability in UTC time links is very important. In [20] we discussed the long-term stability of the SDR linkhere only one example of the Asian link of TL-NICT was analysed and the standard deviation of the DCD between the SDR TWSTFT and GPS PPP links was 1.09 ns. This suggests the SDR link is at least as stable as the GPS PPP link. To complete the study and to have a more general conclusion, we give here other examples of the Europe-Europe-US UTC links over 6 months.

4.3.1 The OP-PTB baseline

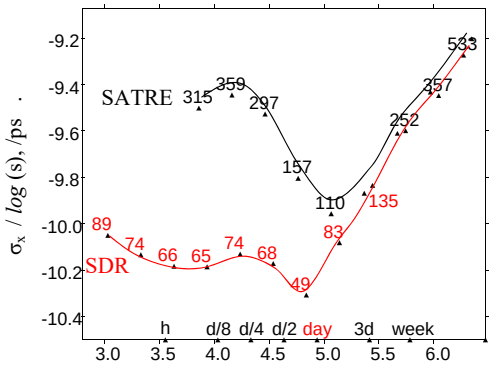


(a) SDR TWSTFT difference



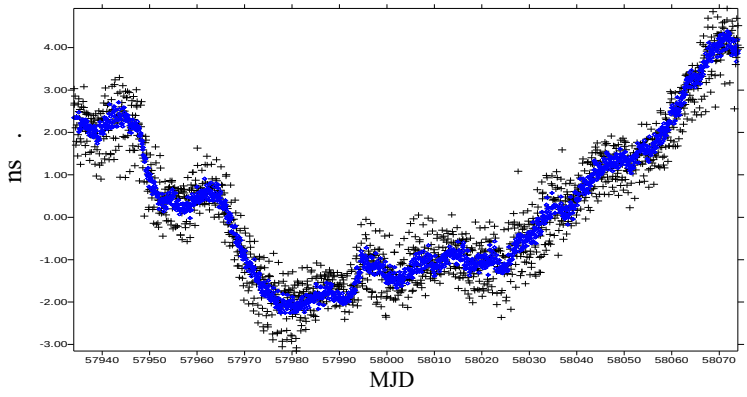
(b) SATRE TWSTFT difference

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

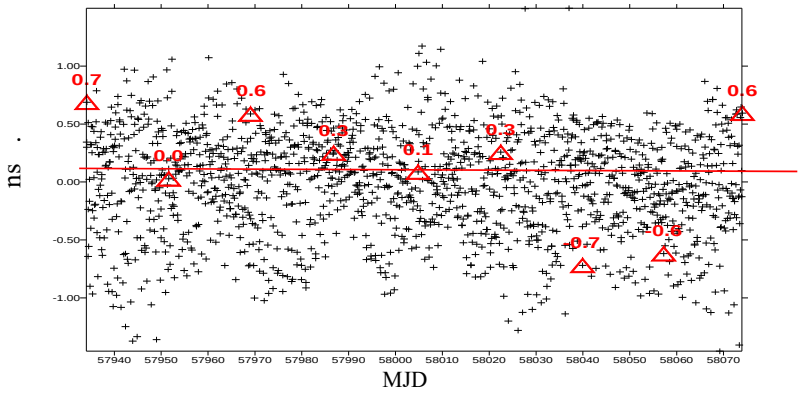


(c) TDev of SDR and SATRE TWSTFT differences

Figure 4.3.1-1 TW SDR and TW SATRE links over the baseline OP-PTB

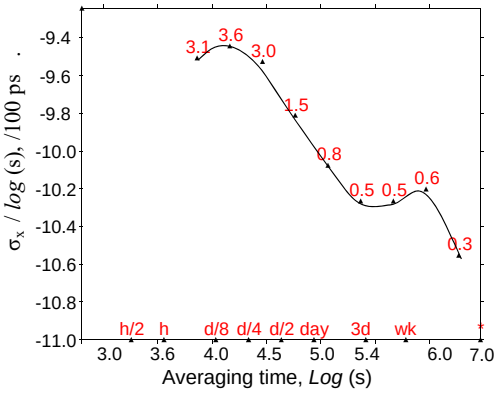


(a) TWSTFT differences of SDR (blue) vs. SATRE (black) links



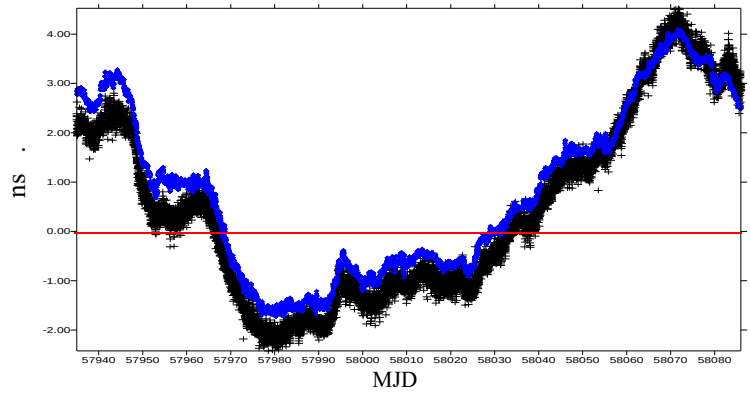
(b) DCD of SDR and SATRE links with the σ(DCD)=0.445 ns with no slope observed

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

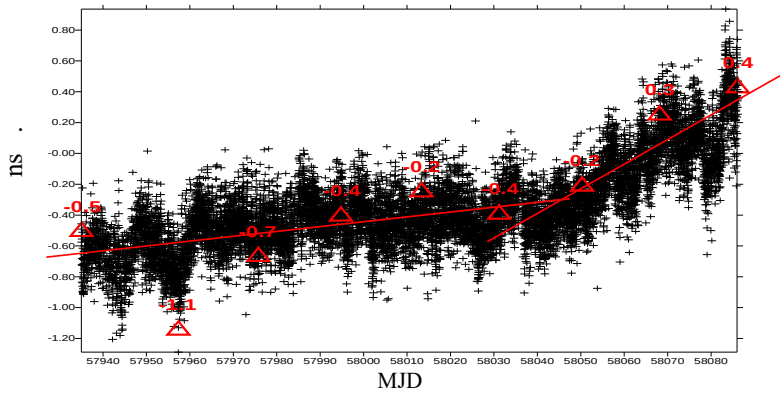


(c) TDev of the DCD of which the instability comes mainly from the noise and diurnal in SATRE measurements

Figure 4.3.1-2 Comparisons of TW SDR and TW SATRE time links over the baseline OP-PTB

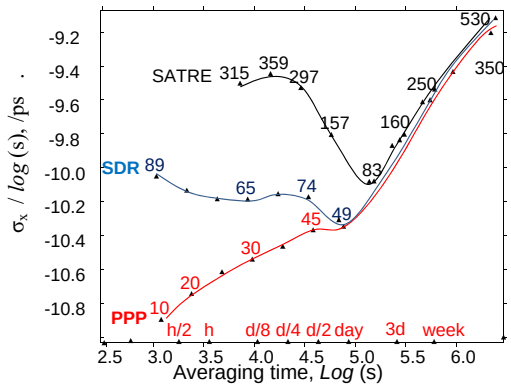


(a) Time transfer difference of SDR (black) and. GPS PPP (blue) links



(b) DCD of SDR and GPS PPP links with the σ(DCD)=0.283 ns and two slopes

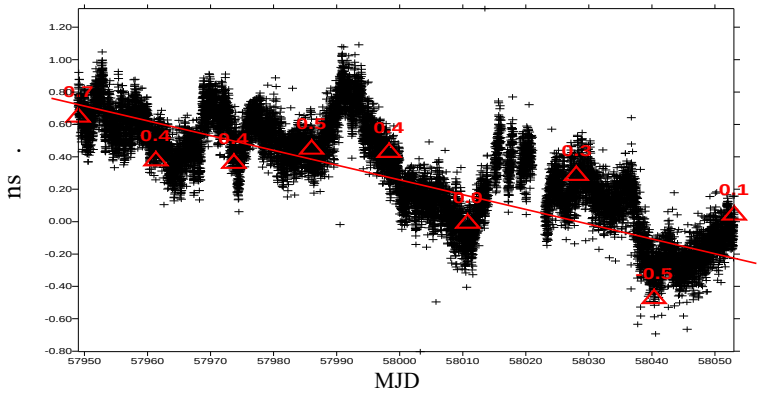
Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.



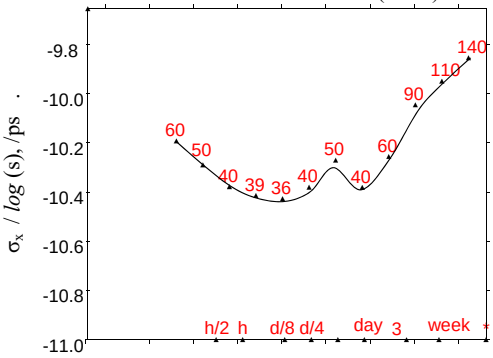
(c) Time deviation of the DCD

Figure 4.3.1-3 The TDev of TW SDR, TW SATRE and GPS PPP over the baseline OP-PTB

4.3.2 The baseline TL-KRIS



(a) DCD of SDT and GPS PPP links with the σ(DCD)=0.314 ns and a slope



(b) Time deviation of the DCD

Figure 4.3.2-1 Comparison between SDR and GPS PPP time links over the baseline TL-KRIS

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

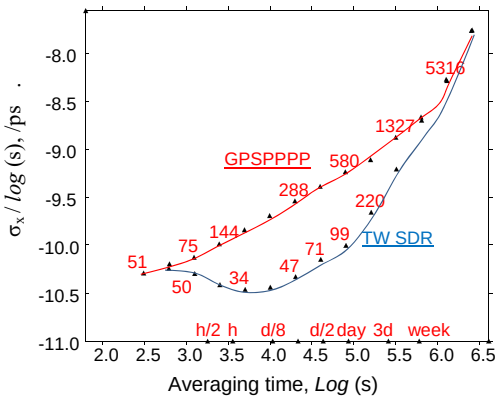


Figure 4.3.2-2 Comparison of the TDev of the SDR (blue) and GPS PPP (red) time links over the baseline TL-KRIS

As given in the plot, the TDev reaches to 34 ps on two hours averaging time. The SDR link is more stable than that of the GPS PPP in both short and long-terms.

4.3.3 The baseline CH-PTB

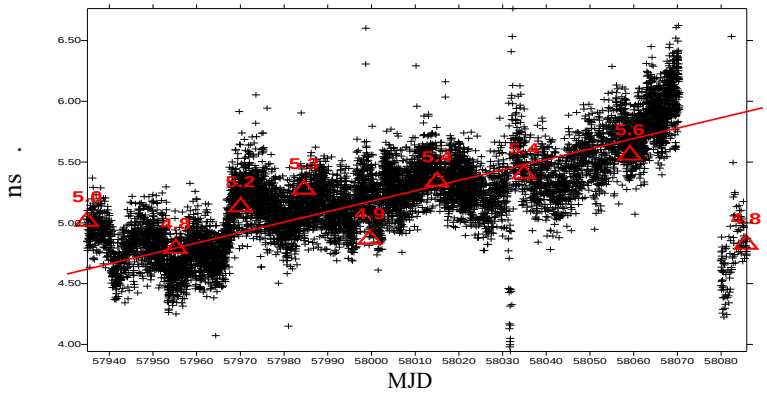


Figure 4.3.3-1 DCD of TW SDR and GPS PPP time links over the baseline CH-PTB with the σ(DCD)=0.389 ns and a slope plus a step in the DCD

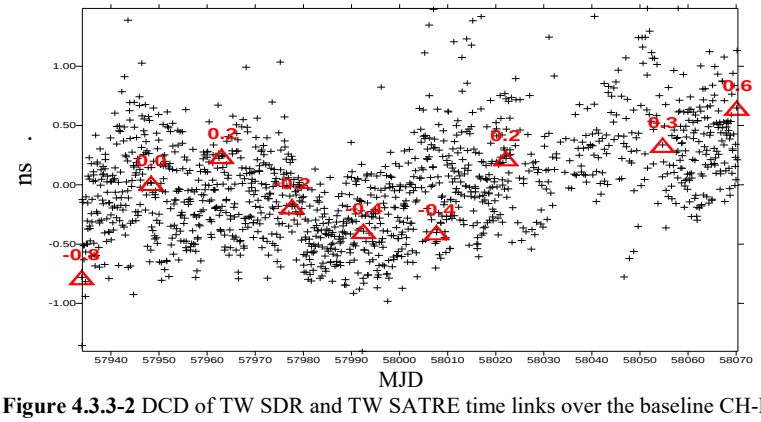


Figure 4.3.3-2 DCD of TW SDR and TW SATRE time links over the baseline CH-PTB with the σ(DCD)=0.441 ns without slop observed

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

4.3.4 The baseline IT-PTB

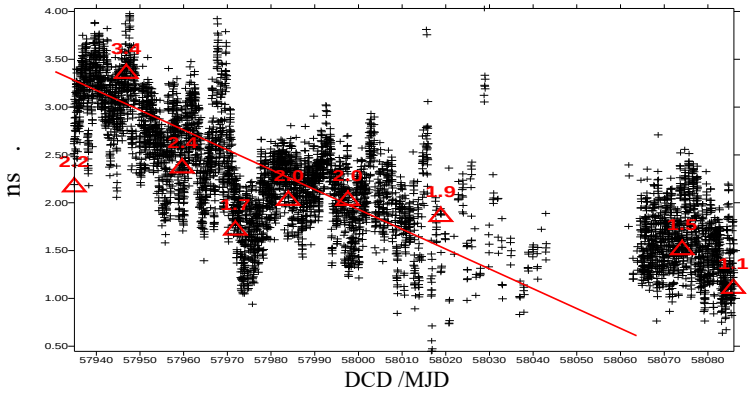
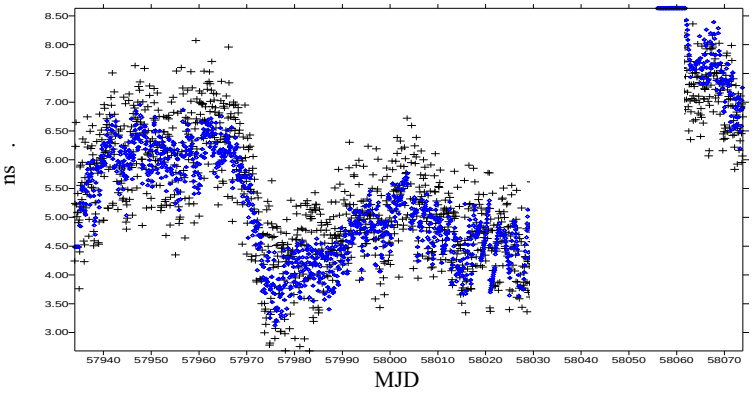
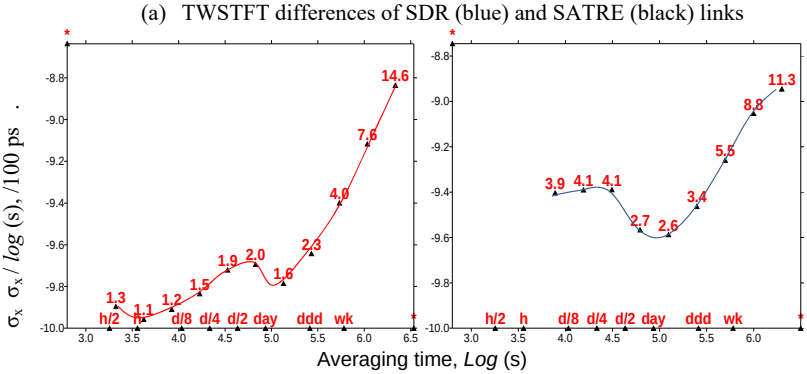


Figure 4.3.4-1 DCD of SDR and GPS PPP time links over the baseline IT-PTB with the σ(DCD)=0.653 ns and a slop



(a) TWSTFT differences of SDR (blue) and SATRE (black) links



(b) TDev of SDR (left) and SATRE (right)TWSTFT differences

Figure 4.3.4-2 Comparison between TW SDR and SATRE time links over the baseline IT-PTB

Here, the up plot is the DCD of the two links with the σ(DCD)=0.708 ns without slope. The down plot is the TDev of the two links. Again, as shown in the Figure, the stability of SDR reaches 110 ps in about one hour and is much more stable than that of the SATRE. Between the 58030 and 58060, both the TW SDR and SATRE data were disturbed and not shown here.

4.4 Summary of the 6-month analysis

Based on the analysis above, we have the following observations:

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

- There are time steps and gaps, in particular, in the links of IT-PTB, AOS-PTB and NIST-PTB. The results of AOS-PTB and NIST-PTB can be found in TM273 [23];
- IT-PTB link is a little nosier than normal;
- Similar conclusions as given in the earlier study, c.f. the TM267 [6] can be made:
  - For the inner-continental links, the TDev of the SDR links is greatly improved as compared to that of the SATRE links in both the measurement noise and the diurnal;
  - For the link NIST-PTB, the measurement noise is improved but the diurnal, which is already very small and in fact the smallest of all the UTC links, is not considerably improved. In fact, the SATRE TWSTFT link NIST-PTB is the most stable one in the TW network. The short term TDev or u_A is about 100 ps, much smaller than the conventional u_A of 200 ps, see section 2.2. That is, 200 ps is almost the uncertainty limit of SDR (for the satellite T11N). It is hence difficult to further improve the uncertainty in an order of 100 ps. In Asia, it seems the links with the satellite 172B are more stable;
  - The best operated and the most stable SDR link is of OP-PTB which can be readily used as a UTC link
- At present, not all the SDR ITU data are as reliable (in terms of data gaps and time steps) as that of SATRE. The data quality will be improved with more experiences in operating and monitoring SDR TWSTFT. A coordinator is to be named to monitor the SDR measurement quality and find a trouble-shooting solution when necessary.

We make here a summary of the 1-month and 6-month analyses:

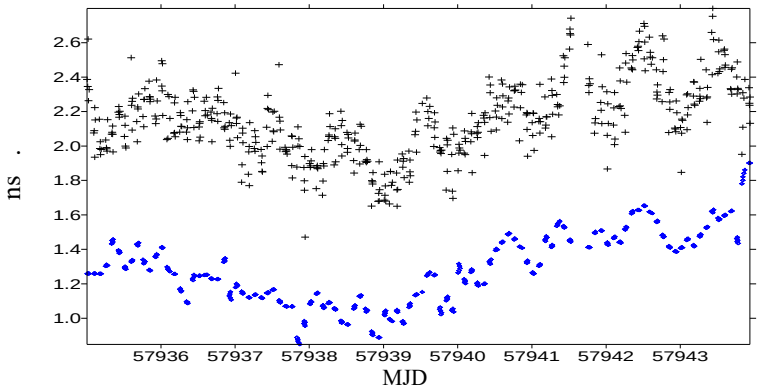
- We confirm the earlier conclusions: the diurnals are considerably reduced, in particular for the inner-Europe links. In consequence, the statistical uncertainty is decreased by a factor of 2 or 3 in most cases;
- In long-term and in normal case (without steps due to the changes in hardware and the software), the SDR links are consistent with GPS PPP links, e.g., for the period of 183 day, $\sigma(DCD_{SDR\text{-}GPSPPP})$=0.283 ns for the baseline of OP-PTB; $\sigma(DCD_{SDR\text{-}GPSPPP})$=0.314 ns for the baseline of TL-KRIS; $\sigma(DCD_{SDR\text{-}GPSPPP})$=0.389 ns for the baseline of CH-PTB and $\sigma(DCD_{SDR\text{-}GPS\ PPP})$=0.653 ns for the baseline of IT-PTB. AOS and NIST suffered several huge steps and we could not make the long-term comparisons;
- In long-term the SDR and SATRE links are well consistent with each other;
- However, we observe a kind of slope between the SDR and GPS PPP links, which seems in the same order as that between SATRE and GPS PPP links. Further studies will be made in this direction;
- The major issue is the steps in the SDR measurement data. The operators of the SDR TWSTFT should pay more attention to the SDR operation at least at the beginning of the SDR TWSTFT operation.

5. STEPS TO FURTHER IMPROVE THE SDR QUALITY

We have pointed out that the uncertainty of the SDR link may be further improved by:

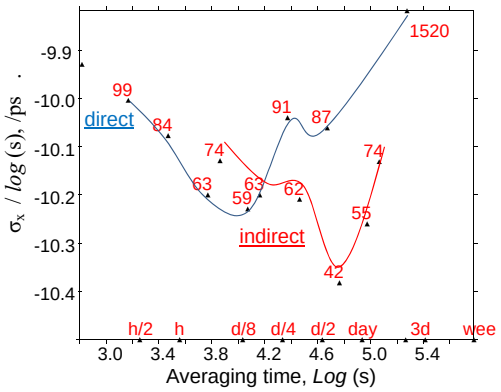
- The indirect link method [10,11] as a simplified application of the TW network time transfer [8];
- The combination of the TW SDR and the GPS PPP which contains the carrier phase information, namely SDR+PPP [14].

As we also know, the GPS IPPP is at present the most precise link available [21] among the techniques used in the practice of the clock comparison. We may therefore use the IPPP to investigate the gains of the improved results obtained by the above two methods. In the followings, we show examples of indirect SDR link, SDR+PPP link and comparing them to IPPP link over the OP-PTB baseline.



(a) TWSTFT differences of Indirect SDR (blue) vs. direct SDR (black) links with the σ(DCD)=0.159 ns

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.



(b) TDev of the TWSTFT differences for direct and indirect SDR links

Figure 5.1 Comparison between direct and indirect (via NIST) TW SDR links over the baseline OP-PTB

The time stability curves show clearly that the TDev of the indirect link is much more stable than that of the direct link. Further investigation beyond the physical reasons may refer to [10, 11] and also [18].

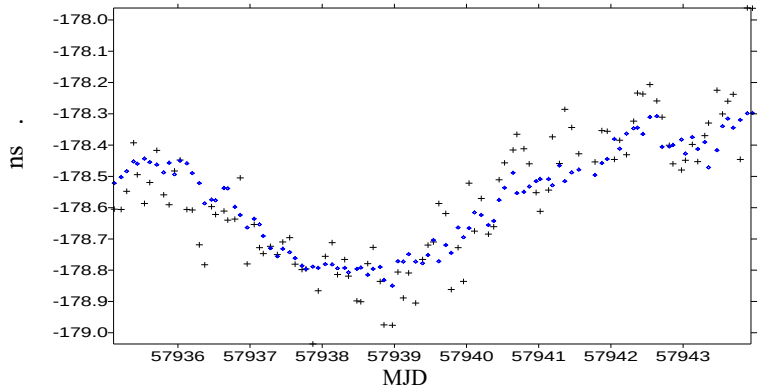
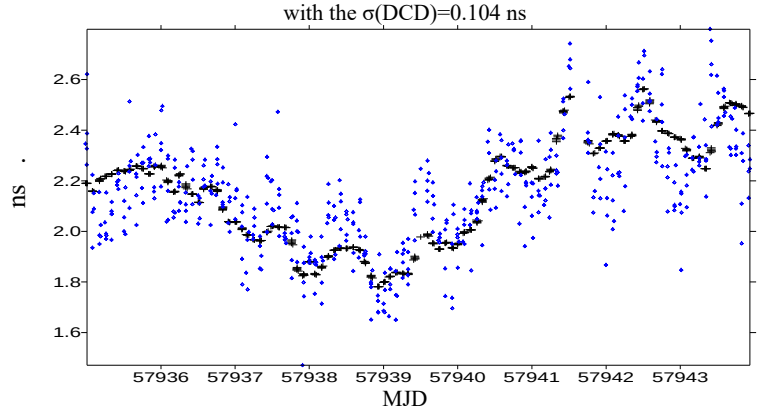
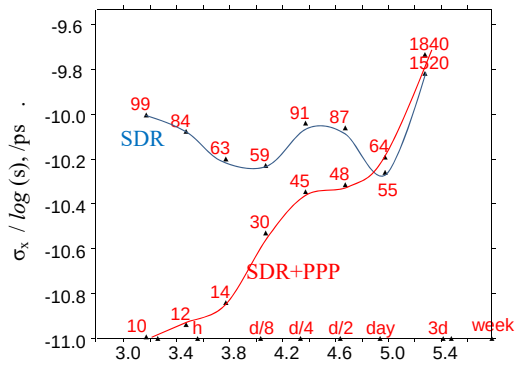


Figure 5.2 Comparison between GPS IPPP (blue) and TW SDR indirect (black, via NIST) links over the baseline OP-PTB with the σ(DCD)=0.104 ns



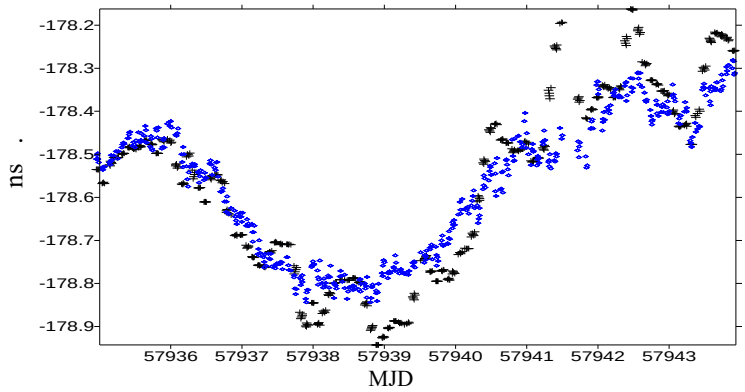
(a) Time transfer differences of SDR+PPP (black) vs. SDR(blue) links with the σ(DCD)=0.143 ns

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

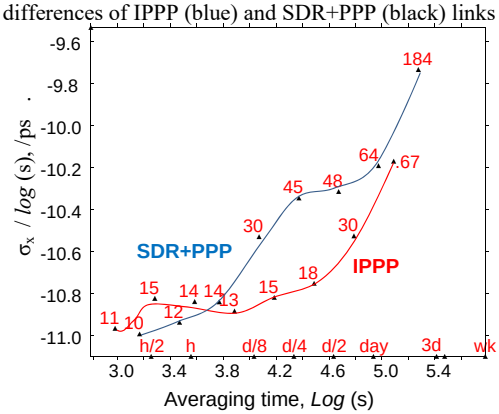


(b) TDev of the SDR+PPP (red) and SDR (blue) links

Figure 5.3 Comparison between the combined TW SDR+GPS PPP and the TW SDR links over the baseline OP-PTB. An excellent stability in the combined SDR+PPP link can be observed



(a) Time transfer differences of IPPP (blue) and SDR+PPP (black) links with the σ(DCD)=0.078 ns

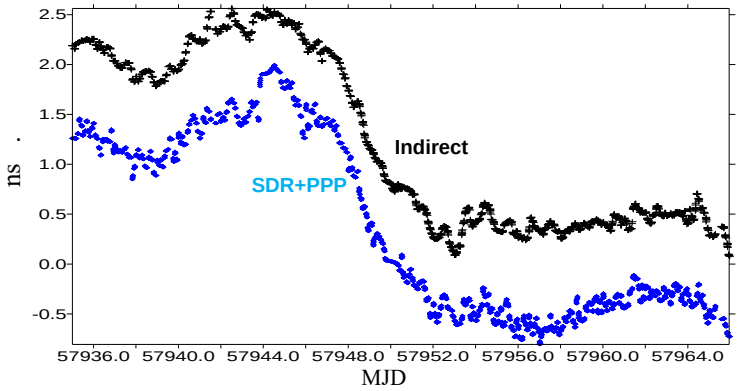


(b) TDev of IPPP (red) and SDR+PPP (blue) links

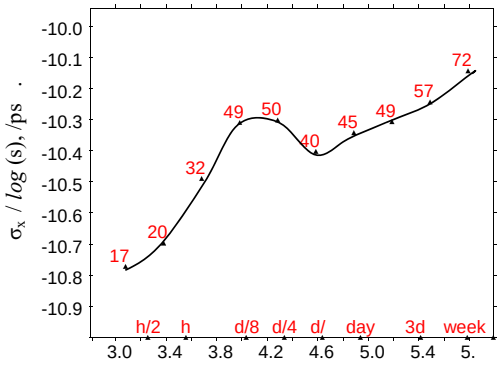
Figure 5.4 Comparison between IPPP and SDR+PPP links over the baseline OP-PTB

In Figure 5.4, we see both IPPP and SDR+PPP links have excellent stability. However, a detailed look shows that the IPPP solution is more stable than the PPP solution in averaging time of 4-5 hours which is about the duration of a passage of a GPS satellite visible by a receiver on the Earth.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.



(a) Time transfer differences of SDR indirect (blue, via NIST) and SDR+PPP (black) links with σ(DCD)=0.112 ns



(b) TDev of the DCD($SDR_{INDIRECT}$-SDR+PPP)

Figure 5.5 Comparison between the TW SDR indirect (via NIST) link and the combined SDR+PPP link over OP-PTB

In the Figure 5.5, it seems that the indirect SDR link is a little noisier than that of the SDR+PPP link. However, it is important to underline that the indirect link is a TWSTFT technique, completely independent from the GPS.

As a summary, table 4.4 takes the IPPP as the reference to compare the other linking techniques: GPS PPP, SDR, SATRE, SDR indirect link via NIST ($SDR_{ind/NIST}$) and the combination (SDR+PPP), cf. the Section 4.2.1 in [23] for the computation details and the plots etc.

Table 4.4 Comparing IPPP to PPP, SATRE, SDR, $SDR_{ind/NIST}$, and the combined link (SDR+PPP) over the baseline OP-PTB

DCD	No	Min/ns	Max/ns	Mean/ns	σ/ns	Gain factor vs. SATRE-IPPP
PPP-IPPP	2880	1.175	1.703	1.378	0.088	-
SDR-IPPP	524	0.198	1.247	0.721	0.158	**3.3**
SATRE-IPPP	118	-1.183	1.638	0.679	0.516	**1.0**
$SDR_{ind/NIST}$-IPPP	105	-0.383	0.201	-0.136	0.104	**5.0**
(SDR+PPP)-IPPP	524	0.547	1.004	0.725	0.078	**6.6***

* Note here that, the solutions (SDR+PPP) and IPPP may not be independent.

In the table, the DCD (double clock differences) is obtained by comparing two links. “Min” and “Max” are the minimum and maximum values of the DCD during the analysed time. Mean and σ are the mean value and standard deviation, respectively. Because the IPPP is the most precise and taken as the reference, the smaller the σ, the more precise the link. The Gain factor in the table is the ratio of the σ of the biggest DCD of SATRE-IPPP 0.516 ns over another link. As can be seen, the most precise link is the combination of SDR and PPP, the gain factor is 6.6. However, SDR+PPP and IPPP are correlated and thus the gain factor should be judged with caution. The second is the indirect SDR link with the gain factor 5.0 and then is 3.3 for the SDR link, that confirms the earlier study. Here we did not compute the Gain factor for PPP-IPPP because they are the same type of measurement using the same raw data set for the computations and therefore strongly correlated.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
”Implementation of SDR TWSTFT in UTC Computation.”
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

Noting that the gain factor of the indirect SDR link $SDR_{ind/NIST}$ vs. the IPPP is 5.0 and the techniques are completely independent from each other. The $SDR_{ind/NIST}$ considerably improves the stability of the SDR, hence it is the most precise TWSTFT technique. Here the NIST in USA is used as the relay station. Because the redundant links, here NIST-OP and NIST-PTB, are always available with the direct link OP-PTB, composing the indirect link does not need to have extra equipment investment or make any extra measurements. Although the indirect TW signals travel distances much longer than the direct ones, the quality is significantly improved, due to the fact that the common affections in NIST-OP and NIST-PTB are mostly cancelled or at least greatly reduced. Not all stations can be used as the relay station of an indirect link, e.g., for the inner-Europe links, none of the European stations can be used as an ideal relay station. They usually increase the instability. The indirect link method works not only for the SDR links but also for the indirect SATRE [10,11]. Further investigations are ongoing.

6. DISCUSSION AND CONCLUSION

In this paper, we discussed the background of the SDR TWSTFT. We quickly reviewed the results of earlier studies and gave the solutions for final open points, including the SDR device setups and the Tsoft software preparation towards this new technique to be used for UTC computation. Technically and practically, the SDR TWSTFT is now ready being accepted for UTC time links. From October 2017, Circular T 359, the SDR link has been used as the backup UTC time link.

The results have been and will continue to be published monthly on the BIPM web site:

ftp://ftp2.bipm.org/pub/tai/timelinks/lkc/1711/opptb/lnk/opptb.tttts.gif for the SDR link;
ftp://ftp2.bipm.org/pub/tai/timelinks/lkc/1711/opptb/dlk/opptb.ttts5.gif for the comparison of SDR-SATRE:
ftp://ftp2.bipm.org/pub/tai/timelinks/lkc/1711/opptb/dlk/opptb.t3sa5.gif for the comparison of SDR-GPS PPP.

As required by the CCTF WG on TWSTFT, the BIPM made the final evaluation using the latest 6-month data from July to December 2017 to better understand the characteristics of the SDR TWSTFT, such as the long-term stability, consistency with existing accurate UTC time transfer links, and find the potential problems in the hardware and the software and fix them.

From the analysis, we summarise our results about the SDR TWSTFT as below:

- We confirm the earlier conclusions, mainly that the diurnals are considerably reduced, in particular for the inner-Europe links. In consequence, the uncertainty is decreased by a factor of 2 or 3 in most cases;
- In long-term and in regular operation (without steps due to the changes in hardware and the software), the SDR links are consistent with the corresponding GPS PPP links, e.g., for the baseline OP-PTB with σ=0.283 ns; TL-KRIS σ=0.314 ns; CH-PTB σ=0.389 ns and IT-PTB σ=0.653 ns. The AOS and NIST SDR links suffered several huge steps and we could not make the long-term comparisons;
- On long-term the SDR and SATRE links are suitably consistent with each other;
- However, we observed a slope in the DCD between the SDR and GPS PPP links, which seems of the same order as that between SATRE and GPS PPP links. If we take into account the slope (the long-term variations between SATRE and GPS PPP links), the σ(DCD) of the SDR and GPS links given above will be much smaller. Further study is to be made in this direction;
- BIPM has made SDR TWSTFT a backup UTC link. It computes the SDR links monthly, makes the link comparisons between the SDR and the other links of SATRE, GPS PPP, GPS IPPP and in the coming future the TWOTFT when possible, thanks to the participations in TWSTFT measurements of AOS and GUM (PL) recently;
- BIPM and the WG on TWSTFT will co-ordinately make decision link by link to introduce the SDR TWSTFT in the UTC computation;
- The major issue is the data gap and time steps in the SDR measurement data, caused mainly by the SDR device maintain and operation. The operators of the SDR TWSTFT operation should pay more attention at least at the beginning of the SDR TWSTFT;

Finally, the conclusion:

The uncertainty of the current (SATRE) TWSTFT is limited by the instabilities (diurnal) of the signal arrival time. The SDR receiver implemented in the TWSTFT earth stations allows a more precise measurement of the arrival time of the coded signal and reduces the diurnal significantly. SDR receiver has been successfully installed at the 15 laboratories AOS, CH, KRIS, IT, NICT, NIM, NIST, NPL, NTSC, OP, PTB, SP, SU, ROA and TL since 2015. Experiments have been carried out continually.

In Jan 2016, a pilot project was launched jointly by BIPM and the CCTF WG on TWSTFT aiming at validating SDR TWSTFT used by laboratories in Asia, Europe and USA with different satellites. The goal of the pilot project is to apply the SDR TWSTFT in UTC generation.

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
"Implementation of SDR TWSTFT in UTC Computation."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

During the last two years, extensive tests and analysis have been made to test the SDR devices, the operational setups, the method and the related software, such as the software packages “Obs2Itu” and the BIPM UTC computation program “Tsoft”.

The gain factors obtained vary from 1.03 to 3.7, and 2 on average, depending on the satellite covering regions and link situations. We can conclude that compared with the SATRE TWSTFT solution, the SDR TWSTFT may improve the time stability by a factor of 2 to 3 on the short-term (in 1-2 hours) and its long-term stability is consistent with the major existing UTC time links such as the TW SATRE and GPS PPP links.

For the Circular T computation, the suggested conventional uncertainties of the TW SDR are:

- u_B equals to the corresponding uncertainty of SATRE link;
- u_A to be 0.2 ns.

Further studies may focus on how to improve the accuracy of the SDR technique by combining it to GPS carrier phase and by fully use of the redundancy in the TWSTFT network, that is, forming the so called ‘indirect’ time links. Encouraging results have been obtained. Another important issue is the TW SDR link calibration, of which the procedure is agreed upon by the WG on TWSTFT and BIPM.

The SDR is an open platform for the operators of the TWSTFT ground stations. People may make deeper studies, for example, by taking out the carrier phase information for various applications.

ACKNOWLEDGMENTS

The authors appreciate for the technical supports from their TWSTFT colleagues and the CCTF Working Group on TWSTFT, especially the supports from TL. We also appreciate NICT for providing us the transponder resource of Eutelsat 172B for the Asia link. We thank specially Drs. Tadahiro Gotoh and Wen-Hung Tseng for their earlier contributions in this study.

†**Disclaimer:** Commercial products are identified for the sake of technical clarity. No endorsement by the BIPM or a pilot project participant is implied. We further caution the readers that none of the described equipment’s apparent strengths or weaknesses may be characteristic of items currently marketed.

REFERENCES

[1] BIPM Circular T 351, April 2017, ftp://ftp2.bipm.org/pub/tai//Circular-T/cirthtm/cirt.351.html
[2] TWSTFT Calibration Guidelines for UTC Time Links V2016 ftp://tai.bipm.org/TFG/TWSTFT-Calibration/Guidelines
[3] Huang Y.J., Tseng W. H., Lin S.Y., Yang S. H. and Fujieda M. (2016) TWSTFT Results by using Software-Defined Receiver Data, Proc. EFTF2016
[4] Jiang Z, Y.J. Huang, W. H. Tseng, S.Y. Lin, M. Fujieda, S. H. Yang and J. Yao (2016) Experience of TWSTFT SDR – BIPM Analysis, BIPM TM260
[5] Reports on SDR from NIST, SU, TL, BIPM etc. 25th Annual Meeting of the CCTF WG on TWSTFT, http://www.bipm.org/wg/CCTF/WGTWSTFT/Restricted/welcome.jsp
[6] JIANG Z and HUANG Y.J. (2017) Status of the pilot study on the TWSTFT SDR for UTC time links, BIPM Technical Memorandum, TM267 2017
[7] ITU Radio communication Sector 2015 The operational use of two-way satellite time and frequency transfer employing PN codes Recommendation ITU-R TF.1153-4 (Geneva, Switzerland).
[8] Jiang Z (2008) “Towards a TWSTFT network time transfer”, Metrologia, IOP Publishing Ltd, vol. 45, pp. 6–11, 2008, doi:10.1088/0026-1394/45/6/S02
[9] Tseng W. H., S. Y. Lin, K. M. Feng, M. Fujieda, and H. Maeno (2010) Improving TWSTFT Short-Term Stability by Network Time Transfer, IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 57, pp. 161-167, 2010
[10] Zhang V, T Parker, S Zhang (2016) A Study on Reducing the Diurnal in the Europe-to-Europe TWSTFT Links, Proc. EFTF2016
[11] Jiang Z, V Zhang, T E Parker, J Yao, Y. Huang and S. Lin (2017) Accurate TWSTFT time transfer with indirect links, Proc. PTTI2017
[12] Tseng W. H., S. Y. Lin, K. M. Feng (2008) Analysis of the Asia-Pacific TWSTFT network, Proc. IFCS, pp. 487-492, 2008

Jiang, Zhiheng; Arias, Felicitas; Zhang, Victor; Huang, Yi-Jiun; Ackhar, Joseph; Piester, Dirk; Lin, Shinn-Yan; Wu, Wenjun; Naumov, Andrey; Yang, Sung-hoon; Nawrocki, Jerzy; Sesia, Ilaria; Schlunegger, Christian; Liang, Kun.
”Implementation of SDR TWSTFT in UTC Computation.”
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

[13] Recommendation of the 2017 CCTF, On improving Two-Way Satellite Time and Frequency Transfer (TWSTFT) for UTC Generation, http://www.bipm.org/en/committees/cc/cctf/publications-cc.html
[14] Jiang Z, Petit G, Combination of TWSTFT and GNSS for accurate UTC time transfer, Metrologia 46 (2009) 305-314
[15] Jiang Z, S. Y. Lin and W. H. Tseng (2017) Fully and optimally use the redundancy in a TWSTFT network for accurate time transfer, Proc. EFTF2017, Besonçon, July 2017
[16] CCTF 2017 Recommendation on the utilization and monitoring of redundant time transfer equipment in the timing laboratories contributing to UTC
[17] Jiang Z, Y Huang, V. Zhang and D Piester (2017) BIPM 2017 TWSTFT SATRE/SDR calibrations for UTC and Non-UTC links, BIPM TM268
[18] V. Zhang, J. Achkar, Y.-J. Huang, Z. Jiang, S.-Y. Lin, T. Parker, D. Piester (2017) A Study on Using SDR Receivers for the Europe-Europe and Transatlantic TWSTFT Links; Proc. 2017 Precise Time and Time Interval Meeting – ION PTTI 2017, 30 Jan – 2 Feb 2017, Monterey, CA, USA, pp. 206-218.
[19] Jiang Z, Czubla A, Nawrocki J, Lewandowski W and Arias F (2015), "Comparing a GPS time link calibration to an optical fibre self-calibration with 200 ps accuracy", 2015 ***Metrologia 52*** 384 doi:10.1088/0026-1394/52/2/384
[20] JIANG Z and Elisa Felicitas ARIAS (2017) Pilot study on the validation of the Software-Defined Radio Receiver for TWSTFT -- BIPM contribution to the validation of the SDR for TWSTFT, Proc. PTTI 2017, January 30-February 2, 2017, Monterey, California, USA
[21] Petit G, A Kanj, S Loyer, J Delporte, F Mercier and F Perosanz (2015), 1×10^{-16} frequency transfer by GPS PPP with integer ambiguity resolution, ***Metrologia 52*** (2015) 301–309
[22] Ad hoc Group (2017) Implementation of SDR TWSTFT in UTC Computation- Road Map of July 2017-Jan. 2018, BIPM TM269
[23] Jiang Z V ZHANG, Y J HUANG, J ACHKAR and D PIESTER (2018) Use of the TWSTFT SDR in UTC Computation -- *Last step of the Road Map for the BIPM-CCTF TWSTFT WG pilot project,* BIPM TM273
[24] M. Fujieda, S-H. Yang, T. Gotoh, S-W. Hwang, H. Hachisu, H. Kim, Y. K. Lee, R. Tabuchi, T. Ido, W-K. Lee, M-S. Heo, C. Y. Park, D-H. Yu, G. Petit (2017), "Advanced Satellite-based Frequency Transfer at the 10^{-16} Level," arXiv:1710.03147
[25] Release V2018.1 of the SDR software package (2018): on the BIPM web site in the folder 'SDR': https://www.bipm.org/wg/CCTF/WGTWSTFT/Restricted/welcome.jsp

AC Voltage Measurements up to 120 V with a Josephson Arbitrary Waveform Synthesizer and an Inductive Voltage Divider

I. Budovsky*, D. Georgakopoulos*, and S. P. Benz†

*National Measurement Institute Australia, Lindfield NSW 2070, Australia
dimitrios.georgakopoulos@measurement.gov.au

† National Institute of Standards and Technology, Boulder, CO 80305, USA

***Abstract* — We use a precision inductive voltage divider and a lock-in amplifier to extend the voltage range of a Josephson arbitrary waveform synthesizer with a maximum voltage of 250 mV rms to provide accurate voltage measurements up to 120 V. Experimental results show that the output of the JAWS can be extended to high voltages with uncertainties of a few microvolts per volt or less, depending on the voltage amplitude and frequency.**

***Index Terms* — Inductive voltage divider, Josephson junction array, measurement standards, measurement techniques, quantum voltage standards.**

I. Introduction

The Josephson arbitrary waveform synthesizer (JAWS) [1] has the potential to replace thermal voltage converters (TVCs) as a primary standard of alternating voltage with the ultimate accuracy of the quantum Josephson effect. However, the voltage range of JAWS technology is presently limited to 1 V rms per Josephson junction array.

At the National Measurement Institute Australia (NMIA) we have developed a 1000 V inductive voltage divider (IVD) of excellent ratio accuracy [2] that can potentially extend the voltage range of the JAWS up to a factor of 1000 without an appreciable increase in the uncertainty. In collaboration with the National Institute of Standards and Technology (NIST), we have been developing such a standard [3]. This paper presents a continuation of this work. We combined a 250 mV NIST JAWS system with the NMIA IVD and a lock-in amplifier and demonstrate agreement with thermal converters at voltages up to 120 V and frequencies up to 1 kHz within the uncertainties of the thermal converters.

II. System Operation

Fig. 1 shows a simplified diagram of the system. A stable ac voltage from a semiconductor-based source is applied to an ac measurement standard (the unit under test) and to the input of the IVD. The ac measurement standard is traceable to primary TVCs. The ac source is synchronized with the JAWS using an arbitrary waveform generator referenced to the same 10 MHz clock as the JAWS. This 10 MHz clock is derived from the cesium primary standard that is designated as the Australian national frequency standard. The phase difference between the output voltages of the IVD and the JAWS can be adjusted within 0.001°, and the difference between these voltages is applied to the lock-in amplifier, which is set to measure the in-phase component of the difference voltage.

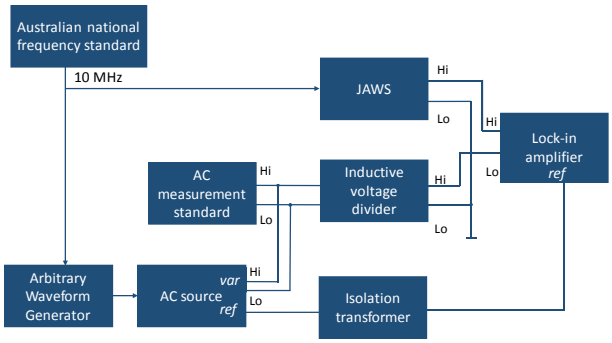


Fig. 1. Block diagram of the measurement system setup

The difference voltage is formed using coaxial cables and a modified tee connector and is applied to a single input of the lock-in amplifier. This is the main improvement from [3] where a differential summation function of the lock-in amplifier was relied upon.

The JAWS system is described in [4] and uses NIST Josephson junctions and microwave circuit designs [5]. The ratio uncertainties of the IVD range from 1×10^{-9} to 7×10^{-7} at frequencies from 40 Hz to 1 kHz. The ac source is a Clarke-Hess§ 5500 phase standard modified to accept an external reference clock whose frequency can be adjusted within a small range to enable synchronization.

Two essential factors for achieving quantized operation of the JAWS are carefully laid out connections and an absence of ground loops. The circuit measurement ground is connected to the mains safety conductor (Earth) at the low potential output of the IVD only. The cases of the ac source and the lock-in amplifier are disconnected from the mains safety conductor to mitigate the effects of common-mode input voltage with respect to the mains safety conductor. The guard of the ac measurement standard is connected to the measurement ground but disconnected from its internal ground terminal. To avoid ground loops through the GPIB computer interface, the lock-in

§ Commercial instruments are identified in this paper only to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Measurement Institute Australia or the National Institute of Standards and Technology, nor does it imply that the equipment identified is necessarily the best for the purpose.

Budovsky, Ilya; Georgakopoulos, Dimitrios; Benz, Samuel.
"AC Voltage Measurements up to 120 V With a Josephson Arbitrary Waveform Synthesizer and an Inductive Voltage Divider."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

amplifier is connected to the control computer through an optically isolated GPIB extender.

The system operation depends on the adjustment of (a) the phase of the voltage between the ac source and the JAWS and (b) the reference phase setting of the lock-in amplifier with respect to its input. To adjust these phases we use the following procedure. First the cable is disconnected from the IVD output which is then shorted at the plug. Then the reference of the lock-in amplifier is adjusted to minimize the quadrature component of the lock-in amplifier reading. Finally the cable is reconnected to the output of the IVD and the phase of the ac source adjusted by, again, minimizing the quadrature component of the lock-in amplifier reading.

III. Experimental Results

The precision evaluation of the JAWS system at voltages greater than 10 V is limited by the increasing uncertainty of the thermal ac-dc transfer standards. For this reason, we first evaluated our JAWS-based system at 7 V rms using a TVC with a low uncertainty (2 μV/V), and then proceeded to 70 V and 120 V. We used the same IVD for both experiments, which, given the very high voltage linearity of the IVD [2], gives us confidence that the type B uncertainty of the system is independent of the voltage amplitude.

In the first experiment, we calibrated a Keysight 3458A configured as a digital sampling voltmeter [6] at 7 V rms and at frequencies from 60 Hz to 1000 Hz using the JAWS and the NMIA conventional TVC-based ac-dc transfer calibration system. The ratio of the IVD was 0.01 and the output voltages of the JAWS and the IVD were 70 mV. The results are shown in Fig. 2. In a second experiment we calibrated a Fluke 5790A ac measurement standard at 70 V and 120 V, and at 60 Hz using the JAWS and the TVC-based ac-dc calibration system (Fig. 3). The IVD ratio in these measurements was 0.001.

IV. Conclusion

In both experiments, the JAWS-based system agreed with the NMIA conventional ac-dc transfer standards within their uncertainties. At low voltage (7 V), the largest difference was 1.3 μV/V, and at high voltage the largest difference was 4 μV/V at 120 V. Further experimental results will be presented at the conference (including different excitation voltages at different frequencies, different output voltages of the IVD for the same excitation voltage obtained with different ratios of the IVD). The uncertainty components will also be discussed in detail.

References

[1] S.P. Benz and C.A. Hamilton, "A pulse-driven programmable Josephson voltage standard," *Appl. Phys. Lett.,* vol. 68, no. 22, pp. 3171 – 3173, May 1996.

[2] G.W. Small, I.F. Budovsky, A.M. Gibbes and J.R. Fiander, "Precision three-stage 1000 V/50 Hz inductive voltage divider," *IEEE Trans. Instrum. Meas.,* vol. 54, no. 2, pp. 600 – 603, April 2005.

[3] T. Hagen, I. Budovsky, S.P. Benz and C.J. Burroughs, "Quantum calibration system for ac measurement standards using inductive voltage dividers," CPEM 2012 Conf. Digest, p. 672, July 2012.

[4] S.P. Benz, C.A. Hamilton, C.J. Burroughs, T.E. Harvey, L.A. Christian and J.X. Przybysz, "Pulse-driven Josephson digital/analog converter," *IEEE Trans. Appl. Supercond.,* vol. 8, no. 2, pp. 42 – 47, June 1998.

[5] S.P. Benz, S.B. Waltman, A.E. Fox, P.D. Dresselhaus, A. Rufenacht, J. M. Underwood, L.A. Howe, R.E. Schwall, and C.J. Burroughs, "One-volt Josephson arbitrary waveform synthesizer," *IEEE Trans. Appl. Supercond.,* vol. 25, no. 1, pp.1 – 8, February 2015.

[6] G. Gubler and E.Z. Shapiro, "Implementation of sampling measurement system for new VNIIM power standard," CPEM 2012 Conf. Digest, p. 294, July 2012.

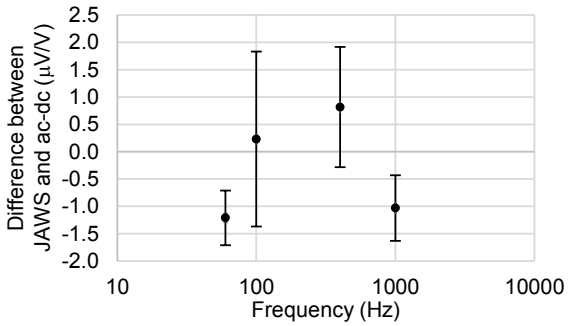


Fig. 2 Difference between the ac-dc difference of a digital sampling voltmeter measured using a TVC and the JAWS-based system at 7 V rms. The error bars show the standard deviation of the JAWS. measurement.

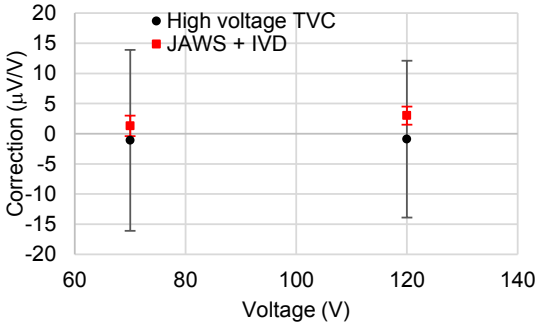


Fig. 3 Ac-dc difference of an ac measurement standard (Fluke 5790A) at 70 V rms and 120 V rms measured with a high voltage TVC and the JAWS-based system. The error bars show the TVC measurement uncertainty and the standard deviation of the JAWS measurement, respectively.

Budovsky, Ilya; Georgakopoulos, Dimitrios; Benz, Samuel.
"AC Voltage Measurements up to 120 V With a Josephson Arbitrary Waveform Synthesizer and an Inductive Voltage Divider."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

Using Temperature to Reduce Noise in Quantum Frequency Conversion

Paulina S. Kuo[1,*], Jason S. Pelc[2], Carsten Langrock[2], and M. M. Fejer[2]
[1]*Information Technology Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899*
[2]*E. L. Ginzton Laboratory, Stanford University, 348 Via Pueblo Mall, Stanford, CA 94305*
[*]*pkuo@nist.gov*

Abstract: A main source of noise in quantum frequency conversion is spontaneous Raman scattering, which can be reduced by lowering the operating temperature. We show reduction in dark count rates that agrees well with theory.
OCIS codes: (190.7220) Upconversion; (190.5650) Raman effect

1. Introduction

Quantum frequency conversion (QFC) will be required in hybrid quantum networks to interface between nodes that operate at different wavelengths and to enable long-distance transport at telecommunications wavelengths. Typically, a QFC device consists of guided-wave $\chi^{(2)}$ device, like a periodically poled lithium niobate (PPLN) waveguide, that performs sum- or difference-frequency generation in the high-conversion regime. System conversion efficiencies up to 45% in single-stage QFC have been demonstrated [1]. Noise photons are mainly due to spontaneous Raman scattering (SRS) and spontaneous parametric downconversion (SPDC) [2]. We focus on SRS noise photons since SPDC noise is not significant in the long-wavelength pumping-scheme employed here [3].

The Raman-scattered photons are associated with the strong pump beam. The PPLN waveguide was designed such that the pump had the longest wavelength to avoid SPDC at the signal wavelength and also have less-efficient anti-Stokes Raman-scattered photons contribute to the noise. Raman scattering is a temperature-dependent process whose efficiency decreases with decreasing temperature. In this work, we describe the theoretically expected relative anti-Stokes Raman noise and show that it agrees well with experimental data.

The spectrum of spontaneous anti-Stokes Raman scattering is given by [3,4]

$$I(\Delta\nu, T) = I_0 \sigma_0(\Delta\nu) n(\Delta\nu, T) \tag{1}$$

where

$$n(\Delta\nu, T) = \left[\exp\left(\frac{hc\Delta\nu}{kT}\right) - 1\right]^{-1}. \tag{2}$$

h is Planck's constant, c is the speed of light, $\Delta\nu$ is the detuning between the wavelength of the Raman scattered photon and the pump, k is Boltzmann constant, and T is the temperature. The frequency-dependent cross-section, $\sigma_0(\Delta\nu)$, describes the Raman spectrum at a fixed temperature. For PPLN, this spectrum is given in Ref. [3]. There are two major peaks in the Raman spectrum of PPLN at $\Delta\nu$ = -260 cm^{-1} and -630 cm^{-1}. Changing the temperature of the waveguide changes the phasematching wavelengths and therefore $\Delta\nu$.

2. Experiment

The experimental setup is shown in Fig. 1. Frequency conversion was performed in a 52-mm-long PPLN reverse-proton-exchanged waveguide that was designed for sum-frequency generation (SFG) between a 1813-nm

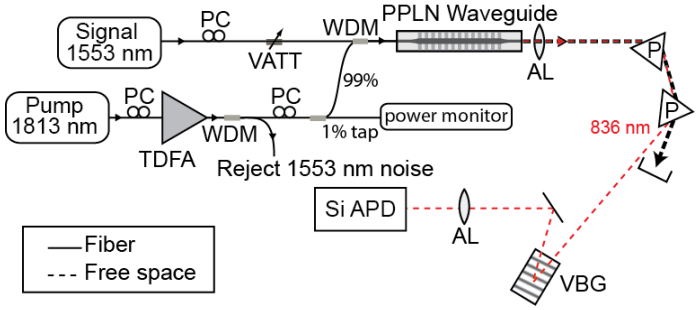


Figure 1. Experimental setup. PC, polarization controller; VATT, variable attenuator; WDM, wavelength division multiplexer; AL, aspheric lens; P, prism; TDFA, thulium-doped fiber amplifer; Si APD, silicon avalanche photodiode; VBG, volume Bragg grating.

Kuo, Paulina; Pelc, Jason; Langrock, Carsten; Fejer, M..
"Using Temperature to Reduce Noise in Quantum Frequency Conversion."
Paper presented at 2018 Conference on Lasers and Electro-Optics: Science and Innovations, San Jose, CA, United States. May 13, 2018 - May 18, 2018.

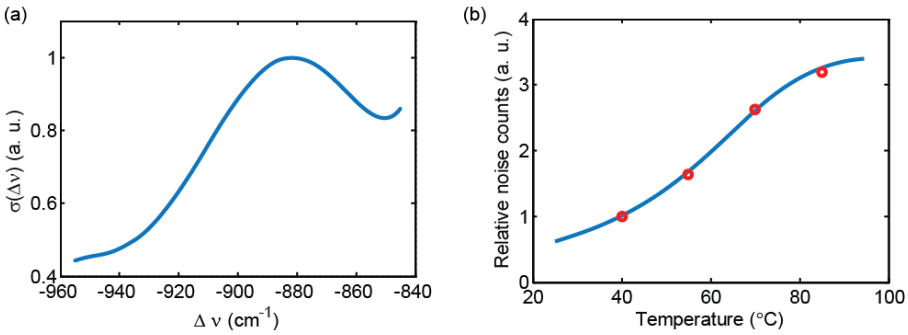


Figure 2. (a) Normalized Raman cross-section near -900 cm^{-1} detuning (after Ref. [3]). (b) Normalized noise counts (relative to T=40 °C). Solid line is theory and red circles are experimental measurements.

Table 1. Temperature tuning results including power at maximum conversion, P_{max}, maximum system conversion efficiency, $\eta(P_{max})$, and dark count rate at maximum conversion, $DCR(P_{max})$. The pump is fixed to 1812.55 nm.

Temperature (°C)	Signal wavelength (nm)	Δν (cm^{-1})	P_{max} (mW)	$\eta(P_{max})$	$DCR(P_{max})$ (counts/sec)
40	1554.14	-917.3	215.4	0.20	2.79e3
55	1557.49	-903.5	207.8	0.19	4.38e3
70	1561.41	-887.4	204.5	0.23	8.74e3
85	1565.57	-870.4	199.9	0.25	11.4e3

pump and a 1553-nm signal at 40 °C; the SFG wavelength is hence at 836 nm. While holding the pump wavelength fixed, we varied the PPLN temperature between 40 °C and 85 °C, which caused the phasematched signal wavelength to tune between 1553 nm and 1565 nm. The input to the waveguide was fiber-pigtailed and the upconverted photons were detected with a Si avalanche photodiode (PerkinElmer SPCM-AQR-14). A prism pair after the waveguide was used to separate residual pump and signal beams from the upconverted photons. We also used a volume Bragg grating (VBG) tunable between 835-840 nm to reduce the Raman noise counts. At near normal incidence, the VBG bandwidth is constant but the diffraction efficiency slightly decreases as the angle of incidence increases (decreasing wavelength). In the analysis, we accounted for the wavelength-dependence of the VBG by normalizing the dark count rates to the system conversion efficiency.

Figure 2a shows the Raman cross-section spectrum for lithium niobate (after Ref. [3]) across the frequency range of interest. Using this data and Eqs. (1) and (2), we calculated the theoretical Raman scattering intensity, shown as the solid line in Fig. 2b (normalized to the value at 40 °C). The circles in Fig. 2b show the experimental noise count rates divided by the maximum conversion efficiency. There is excellent agreement between theory and experiment. The data used to calculate the red circles are shown in Table 1. We see an increase in system conversion efficiency at higher temperature when the VBG is tilted closer towards normal incidence. Non-ideal photon collection can be seen in certain measurements of $\eta(P_{max})$, which is the reason we divided $DCR(P_{max})$ by $\eta(P_{max})$ in Fig. 2b.

3. Conclusion

In conclusion, we have shown that the reduction in noise counts due to temperature during quantum frequency conversion is well-described by the theory of anti-Stokes Raman scattering, which includes accounting for the Raman cross-section spectrum. We observed a 3-fold decrease in noise counts when the PPLN temperature is reduced from 85 °C to 40 °C. This result suggests an interesting path forward to reducing dark count rates in quantum frequency conversion.

References

[1] C. Langrock, E. Diamanti, R. V. Roussev, Y. Yamamoto, and M. M. Fejer, "Highly efficient single-photon detection at communication wavelengths by use of upconversion in reverse-proton-exchanged periodically poled $LiNbO_3$ waveguides," Opt. Lett. **30**, 1725-1727 (2005).
[2] P. S. Kuo, J. S. Pelc, O. Slattery, Y.-S. Kim, M. M. Fejer, and X. Tang, "Reducing Noise in single-photon-level frequency conversion," Opt. Lett. **38**, 1310-1312 (2013).
[3] J. S. Pelc, L. Ma, C. R. Phillips, Q. Zhang, C. Langrock, O. Slattery, X. Tang, and M. M. Fejer, "Long-wavelength-pumped upconversion single-photon detector at 1550 nm: performance and noise analysis," Opt. Express **19**, 21445-21456 (2011).
[4] R. H. Stolen, "Nonlinear Properties of Optical Fibers," in *Optical Fiber Telecommunications*, S. E. Miller and A. G. Chynoweth, eds. (Academic Press, 1979), pp. 125-150.

Kuo, Paulina; Pelc, Jason; Langrock, Carsten; Fejer, M..
"Using Temperature to Reduce Noise in Quantum Frequency Conversion."
Paper presented at 2018 Conference on Lasers and Electro-Optics: Science and Innovations, San Jose, CA, United States. May 13, 2018 - May 18, 2018.

DC Comparison of a Programmable Josephson Voltage Standard and a Josephson Arbitrary Waveform Synthesizer

Alain Rüfenacht, Nathan E. Flowers-Jacobs, Anna E. Fox, Steven B. Waltman, Robert E. Schwall, Charles J. Burroughs, Paul D. Dresselhaus, and Samuel P. Benz

National Institute of Standards and Technology, Boulder, CO 80305, USA
alain.rufenacht@nist.gov

***Abstract* — We present the first dc comparison of a programmable Josephson voltage standards and a pulse-driven Josephson arbitrary waveform synthesizer (JAWS) at 3 V. Both circuits are mounted side-by-side on the cold stage of a cryocooler. The relative agreement achieved was better than 1 part in 10^8. This measurement allowed us to identify systematic errors of the JAWS system. An undesired current injection from the JAWS isolation amplifier into the measurement circuit was responsible for an error voltage of a few nanovolts.**

***Index Terms* — Digital-analog conversion, Josephson arrays, standards, superconducting integrated circuits, voltage measurement.**

I. Introduction

The Josephson Arbitrary Waveform Synthesizer (JAWS) was developed to generate quantum-accurate waveforms with low harmonic distortion for voltage metrology applications [1]. The programmable Josephson voltage standard (PJVS), with an output voltage of 10 V, is currently used by primary standard laboratories for dc calibrations [2]. PJVS systems are also capable of generating stepwise-approximated waveforms at frequencies up to ~1 kHz. Direct comparison of JAWS and PJVS waveforms at 1 V rms and 250 Hz agree to 1 part in 10^8 [3]. Recent JAWS development at NIST has increased the rms output voltage of a single chip to 2 V, enlarging the voltage overlap domain with the PJVS. Combining JAWS and PJVS for the first time in the same cryostat allows us to test the performance of each system. We compared both systems at 3 V dc, without the additional complications associated with ac waveforms. With the new redefinition of the SI, such a dc comparison is an important step to verify the agreement of JAWS systems with the well-established dc primary voltage standards.

II. Measurement Setup

The measurement setup, shown in Fig. 1, consists of a 10 V PJVS circuit and a "1 V + 1 V" JAWS circuit mounted side-by-side on the cold plate of a cryocooler operated at 4.2 K. The measured cooling capacity of the system at this temperature is about 550 mW [4]. The PJVS circuit has a total of 265 156 Josephson junctions (JJs) and is biased at 15 GHz. Details about the PJVS circuit, system, and the 24-channel current source can be found in Ref. [2]. The JAWS circuit has eight arrays, each with 12 810 JJs, capable of generating a dc voltage of 3.051 V

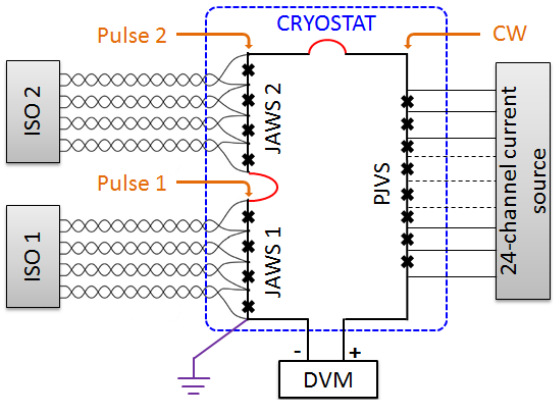


Fig. 1. Schematic of the PJVS and JAWS comparison measurement circuit.

in total when all the JJs are biased with a continuous pulse train generated by a commercial high-speed arbitrary waveform generator clocked at 14.4 GHz. With the "1 V + 1 V" configuration, the two halves of the JAWS circuit (JAWS 1 and JAWS 2) are independent. Each half has a separate pulse bias line (labeled Pulse 1 and Pulse 2 in Fig. 1) connected to two stages of Wilkinson dividers to distribute the pulses among the four arrays [1].

Two short copper wires (red wires in Fig. 1) are soldered between the two JAWS circuits and between the JAWS and PJVS packages, so that all three circuits (JAWS 1, JAWS 2, and PJVS) are connected in series. The JAWS compensation current bias is provided by two custom battery-powered isolation amplifiers (ISO 1 and ISO 2), connected respectively to JAWS 1 and JAWS 2. The voltage difference between the two systems $\Delta V = V_{JAWS} - V_{PJVS}$ is measured by a digital nanovoltmeter (DVM). Copper twisted-pair wires connect the bottom of JAWS 1 array and the bottom of the PJVS array to the DVM at room temperature. The Earth ground potential in the measurement circuit is connected to the bottom of the JAWS 1 array. The fast pulse generator (JAWS) and the PJVS continuous waveform generator (CW) are locked to the 10 MHz frequency from the NIST atomic clock.

III. DC Comparison Results

Before starting the comparison measurement, the leakage current to Earth ground (LCG) from the PJVS and JAWS

Rufenacht, Alain; Flowers-Jacobs, Nathan; Fox, Anna; Waltman, Steven; Schwall, Robert; Dresselhaus, Paul; Benz, Samuel; Burroughs, Charles.
"DC Comparison of a Programmable Josephson Voltage Standard and a Josephson Arbitrary Waveform Synthesizer."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

systems was evaluated with the method described in Ref. [5]. At 3.051 V the LCG measured on the PJVS system was less than 25 pA and the corresponding voltage error on the comparison result was negligible. However, the LCG measured on the JAWS system for the same voltage was much larger (~1.5 µA). The JAWS-PJVS comparison was performed at three different voltages, 3.051 V, 1.526 V, and 0 V, with a total of six different bias configurations of the JAWS system (labeled "A" to "F" in Table 1).

#	PJVS (V)	JAWS 1 (V)	JAWS 2 (V)	ΔV (nV)	**ΔV meas. (nV)**	σ (nV)
A	3.051526899	1.525...	1.525...	**–0.3**	**0.9**	1.0
B	1.525763449	1.525...	0	**0.3**	**–5.0**	1.0
C	1.525763449	1.525...	N.C.	**0.3**	**0.5**	1.0
D	1.525763449	0	1.525...	**0.3**	**7.8**	1.0
E	1.525763449	N.C.	1.525...	**0.3**	**7.8**	1.1
F	0	0	0	**0**	**0.2**	0.8

Table 1. DC voltage comparison results as a function of the JAWS bias configuration. The voltage produced by JAWS 1 and JAWS 2 was 1.525 763 449 35 V @ 14.4 GHz or 0 V. The columns "ΔV" and "ΔV meas." are the expected (calculated) and measured voltage differences between the two systems. The Type-A uncertainty reported is the standard deviation σ with $k = 1$. The abbreviation N.C. indicates that the compensation module (ISO) was disconnected from the array (0 V).

The CW frequency of the PJVS was adjusted to match the voltage of the JAWS system to the 9th decimal place. Small voltage differences, $|\Delta V| < 0.4$ nV, remained. Each value reported for the voltage difference was calculated with a linear fit based on four polarity reversal sets "+–+– ", to remove the contributions of the thermal electromotive forces. A polarity set consisted of 15 DVM readings on the 1 mV range at 10 power line cycles each. To test the quantum locking range of both systems during the comparison measurements, a dither current of ±0.25 mA was sequentially applied to the PJVS and the JAWS arrays. None of the measurement results reported in Table 1 were affected by the applied dither current.

At 3.051 V, the difference between the measured and expected values ΔV was 1.2 nV (Fig. 2). However, a larger spread in the results was obtained at 1.526 V when one of the JAWS array was set to 0 V (with or without the corresponding ISO connected). This effect cannot be explained solely by the JAWS LCG. An independent measurement showed the presence of a significant current injection (CI) from both isolation amplifiers (several microamperes, depending on the potential difference with Earth ground). Any current flowing in the resistive wire connecting the JAWS 1 and JAWS 2 arrays will lead to a voltage error in the measurement circuit. However, when the ISO 2 was disconnected (bias configuration "C") or when the voltages of all the arrays was set to zero ("F"), no current was flowing in the resistive wire. Connecting or disconnecting the ISO 1 ("D" and "E") results in the same voltage error (7.5 nV), which showed that the CI from the ISO 1 unit, when its low side was referenced to the Earth ground potential, did not contribute to the voltage error. Replacing the

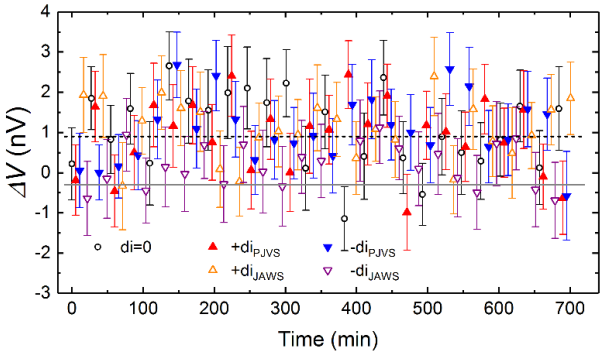


Fig. 2. Measured voltage difference versus time for the comparison at 3.051 526 899 V. The dashed line shows the mean value measured (0.9 nV) while the solid line represents the expected voltage difference (–0.3 nV). Error bars are calculated from the residuals of the fit to remove the thermal electromotive forces (k=1).

"1 V + 1 V" JAWS wiring configuration with the "2 V" configuration that has an on-chip superconducting series connection should eliminate this source of voltage error in future comparisons.

IV. Conclusion

These dc comparison measurements revealed the presence of undesired CI generated by the present ISO. Our present priority is to reduce the LCG and CI caused by the ISO units. After we remove the undesired voltage error due to the CI, the agreement between the JAWS and PJVS systems is expected to improve from 1 part in 10^8 to 1 part in 10^9. Our measurement noise was only $\sigma \cong 1$ nV, which demonstrates that the two very different types of voltages standards can operate side-by-side on a single cryocooler without interference. The next measurements performed with this setup will compare PJVS stepwise-approximated waveforms with spectrally-pure low-frequency JAWS sine waves using the differential sampling method.

References

[1] N. E. Flowers-Jacobs, S. B. Waltman, A. E. Fox, P. D. Dresselhaus and S. P. Benz, "Josephson arbitrary waveform synthesizer with two layers of Wilkinson dividers and an FIR filter," *IEEE Trans. Appl. Supercond*., vol. 26, no. 6, 143007, Sep. 2016.

[2] C. J. Burroughs, P. D. Dresselhaus, A. Rüfenacht, D. Olaya, M. M. Elsbury, Y. Tang and S. P. Benz, "NIST 10 V programmable Josephson voltage standard system," *IEEE Trans. Instrum. Meas.*, vol. 60, no 7, pp. 2482–2488, Jul. 2011.

[3] R. Behr, O. Kieler, J. Lee, S. Bauer, L. Palafox, and J. Kohlmann, "Direct comparison of a 1 V Josephson arbitrary waveform synthesizer and an ac quantum voltmeter," *Metrologia*, vol. 52, no. 4, pp. 528–537, Aug. 2015.

[4] A. Rüfenacht, L. A. Howe, A. E. Fox, R. E. Schwall, P. D. Dresselhaus, C. J. Burroughs, and S. P. Benz, "Cryocooled 10 V programmable Josephson voltage standard," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 6, pp. 1477–1482, Jun. 2015.

[5] S. Solve, A. Rüfenacht, C. J. Burroughs and S. P. Benz, "Direct comparison of two NIST PJVS systems at 10 V," *Metrologia*, vol. 50, no. 5, pp. 441–451, Oct. 2013.

Rufenacht, Alain; Flowers-Jacobs, Nathan; Fox, Anna; Waltman, Steven; Schwall, Robert; Dresselhaus, Paul; Benz, Samuel; Burroughs, Charles.
"DC Comparison of a Programmable Josephson Voltage Standard and a Josephson Arbitrary Waveform Synthesizer."
Paper presented at 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France. July 8, 2018 - July 13, 2018.

User Context: An Explanatory Variable in Phishing Susceptibility

Kristen K. Greene
National Institute of Standards and Technology
kristen.greene@nist.gov

Michelle P. Steves
National Institute of Standards and Technology
michelle.steves@nist.gov

Mary F. Theofanos
National Institute of Standards and Technology
mary.theofanos@nist.gov

Jennifer Kostick
National Institute of Standards and Technology
jennifer.kostick@nist.gov

***Abstract*—Extensive research has been performed to examine the effectiveness of phishing defenses, but much of this research was performed in laboratory settings. In contrast, this work presents 4.5 years of workplace-situated, embedded phishing email training exercise data, focusing on the last three phishing exercises with participant feedback. The sample was an operating unit consisting of approximately 70 staff members within a U.S. government research institution. A multiple methods assessment approach revealed that the individual's work context is the lens through which email cues are interpreted. Not only do clickers and non-clickers attend to different cues, they interpret the same cues differently depending on the alignment of the user's work context and the premise of the phishing email. Clickers were concerned over consequences arising from *not* clicking, such as failing to be responsive. In contrast, non-clickers were concerned with consequences from clicking, such as downloading malware. This finding firmly identifies the alignment of user context and the phishing attack premise as a significant explanatory factor in phishing susceptibility. We present additional findings that have actionable operational security implications. The long-term, embedded and ecologically valid conditions surrounding these phishing exercises provided the crucial elements necessary for these findings to surface and be confirmed.**

Keywords—decision-making, embedded phishing awareness training, user-centered approach, survey instrument, long-term assessment, operational data, trial deployment, network security, security defenses

I. Introduction

The problem of phishing is not solved. It is an escalating cyber threat facing organizations of all types and sizes, including industry, academia, and government [1], [11], [20], [22]. Often using email, phishing is an attempt by a malicious actor posing as trustworthy to install malware or steal sensitive information for financial gain. The nature of phishing itself has changed, moving far beyond "traditional" phishing for usernames, passwords, and credit card numbers via fraudulent websites, and into more sophisticated cybercrime attacks that mount advanced persistent threats against organizations and steal individuals' financial identities with devastating consequences for both users and organizations. The practice of phishing has turned a pervasive means of communication—email—into a dangerous threat channel. Symantec reports that malicious emails were the weapon of choice for bad actors, ranging from state-sponsored espionage groups to mass-mailing ransomware gangs, and that one in 131 emails sent during 2016 was malicious [22].

Workshop on Usable Security (USEC) 2018
18 February 2018, San Diego, CA, USA
ISBN 1-891562-53-3
https://dx.doi.org/10.14722/usec.2018.23016
www.ndss-symposium.org

Advanced threats via ransomware increased 167 times from four million attempts in 2015 to 638 million attempts in 2016, mostly through phishing campaigns [20]. Email's popularity with attackers is driven by several factors. It is a proven attack vector. It does not rely on system vulnerabilities, but on human deception. Routine business processes, such as correspondence about delivery notifications and invoices, provide camouflage for these malicious emails and were the favored guise for spreading ransomware in 2016 [22]. In the information security domain, the use of deception to manipulate individuals for fraudulent purposes is referred to as social engineering.

To help combat the phishing threat, many organizations utilize some type of phishing awareness training to make employees and students more aware of phishing threats and consequences, e.g., Stanford University's Phishing Awareness Service [21]. These embedded phishing awareness training systems use software to send simulated phishing emails to users' regular email accounts. By "phishing" users in their normal computing environments, these emails are intended to train people to recognize and avoid falling victim to phishing attacks in their work (or school) setting. Emails are designed to emulate real-world threats currently facing organizations, providing a realistic experience in a safe, controlled way so recipients can become familiar with the types of tactics used in real phishing attacks. Embedded training schemes that combine training people in their normal work environments with immediate feedback produce more lasting change to behaviors and attitudes [14]. It also provides the means to capture click decisions in an operational environment.

The goal of this work was to better understand why users click and do not click on links or open attachments in phishing training emails that were part of embedded phishing awareness exercises at the National Institute of Standards and Technology (NIST), a U.S. government research institution. From mid-2012 through 2015, the institute's Information Technology Security and Networking Division (ITSND) facilitated 12 operationally-situated exercises using a commercially-available phishing awareness training system. All exercises were conducted in one particular operating unit (OU) at the institute. Although staff knew their OU was participating in these exercises, they were not announced and were conducted at irregular intervals to avoid priming effects. Unfortunately, click rates were variable across the initial 3.5 years of exercises, making the training effect difficult to characterize. In 2016, human factors researchers partnered with the institute's ITSND to better understand the variability in the operational click-rate data. Over the course of 2016, another three exercises were conducted exactly as before with the following exception: each exercise also had an accompanying post-exercise survey to better understand why

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

users were clicking or not clicking. The long-term data from the 2016 exercises alone represents real-world phishing data with breadth and depth unlike any reported in the literature to-date. For all exercises in the 4.5-year span (2012 through 2016), the phishing training emails modeled real-world phishing campaigns, and participating staff were in their normal work environments with their regular work loads, providing ecological validity. This paper presents the novel finding that the alignment of user context and the phishing message premise is a significant explanatory factor in phishing susceptibility, impacting depth of processing and concern over consequences. We believe that the rare opportunity to collect data surrounding phishing click decisions in the workplace coupled with ecologically valid conditions over time provided the crucial elements for these findings to surface. Because of the scarcity of operationally-situated data, we also report on findings we believe to be actionable now in operational settings.

II. Background

A. Background research

Technological and human-centered approaches are used to combat email phishing. Technology-based solutions generally focus on reducing software vulnerabilities, for example, maintaining software currency and identifying malicious websites and phishing emails based on their characteristics. This identification centers around using server-side filtering and classifiers and client-side filtering tools [2]. The server-side mechanisms strive to remove malicious messages and website links before the user sees them, while the client-side tools often attempt to aid user decision-making [9], [11]. Much of the filter algorithm development and classifier training is done offline, prone to error, and reactive in nature [2].

Human-centered approaches attempt to bridge the gap left by reactive technological solutions [9]. Research in these approaches often falls into one of three categories: educational awareness training to identify phishing messages, new user interface mechanisms and designs coupled with client-side filtering intended to aid users' click decisions, and research that considers psychological factors in decision-making with respect to phishing [18].

While there are many studies that have explored the ability of email users to recognize phishing cues, most of these have been conducted in laboratory settings where users are not faced with click decisions in real-world settings under their normal workloads and time pressures, and without laboratory priming effects. In contrast to laboratory-based studies, our user-centered phishing assessment situates participants in the intended use setting: the employee workplace.

There are only a few studies in the literature using embedded, simulated phishing in the user's normal computing environment during their normal work day: [4], [8], [16], and [17]. These studies were set in the real world using operationally-situated study settings, but were focused on investigating training materials and click rates, not on participant click-decision factors. The studies described in [4] and [16] primarily looked at the efficacy of embedded awareness training. Those reported in [8] and [17] focused on the startling number of users who clicked in their respective operational environments. Click decision exploration was not included in [16] and [17], and only speculated about in [8].

The workplace-situated study reported in [4] is the most similar study to our work of those cited, although the focus of that study was to examine the effect of training materials on click decisions rather than other factors surrounding these decisions. Caputo et al. describe three phishing training exercises over the course of eight months. Only after the third exercise was completed were interviews conducted with 27 participants. Although cursory thematic data were reported, details about click decisions were not given, with the caveat that almost all interviewees who clicked most often recalled the third trial only, while most non-clickers did not recall any of the training phish due to the length of time that elapsed between the initial phish and the interview. In contrast, our work centers on investigating factors surrounding participant click decisions rather than training material effectiveness.

There are many reasons why there are so few studies set in the real world. [4], [8], and [16], among others, note challenges in conducting real-world phishing studies that can provide more ecological validity and richer data than those administered in laboratory settings. For example, coordinating with operational staff to get training phish through corporate firewalls and filters are hurdles that must be overcome. Maintaining participant privacy and avoiding participant cross-contamination, such as warning other participants [15], are note-worthy challenges. The expense of examining training retention for longer than a 90-day period [4] and obtaining stakeholder buy-in [8] are also mentioned. Given these significant challenges, why attempt to study phishing click decisions *in the workplace?* The answer centers on *context of use*–how, where, and under what circumstances email users are making click decisions. Indeed, Wang, et al. note the enormous contribution data from real phishing victims would provide, if it were available [27]. The data presented here provide rare insights into click decisions in the workplace.

B. Project background

This project was started in 2012 by the institute's ITSND as a long-term trial deployment of an embedded phishing awareness training effort. The trial deployment was intended as a multi-year effort and used a commercially-available system to help develop and deliver phish messages and training, as well as track click rates. The same system was used throughout the entire 4.5-year period. For all exercises, the targeted population within the institute was one operating unit having approximately 70 staff members. The awareness training provided by these exercises augmented the IT security awareness training the entire institute's workforce received annually. OU staff were aware their unit was participating in the trial. However, exercises were unannounced and deployed at irregular intervals to avoid priming effects.

All exercises were conducted by the OU's Information Technology Security Officer (ITSO). The same person held the position during the entire 4.5-year period. The ITSO selected the phishing message and its premise from templates provided by the training system that mimicked current real-world threats. The ITSO used some messages without modification so response rates could be compared with other organizations using the same training system and message, a form of benchmarking. Some messages were tailored to align with business and communication practices within the organization or were personalized, in other words, they were spear phish [25], [27].

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

The ITSO also developed the training given to those who clicked. Further, the ITSO set the timing of each exercise and coordinated with the IT security team to allow the phish through the institute's firewalls and filtering mechanisms.

Fig. 1 shows the click rates for the exercises conducted in 2012 through 2016. For the first 12 exercises, those without survey feedback, click rates ranged from 1.6 % to 49.3 %, having a *Mean* = 17.3 %, *Median* = 11.9 %, and *SD* = 15.2. The social engineering premise for each exercise was varied, except for the three 'Package' exercises that used the same message mimicking a package delivery notice. This benchmark phish was not tailored to the organization. Click rates for the 'Package' exercises were 28.6 %, 12.2 %, and 7.7 %. Interestingly, despite being seen three times, the click rate for this phish exercise did not go to zero, although it did decline. All premises used imitations of normal discourse [3] and were based on then-current social engineering scams. The sample was the OU; the sole inclusion criterion was being assigned to the OU. The only OU staff member excluded was the OU's ITSO, who conducted the exercises. The OU had an annualized separation rate of 8.7 % during the entire 4.5-year period. Individuals were intentionally not tracked across exercises to protect employee privacy.

Although click rates declined on average during the first 3.5 years of training, the trend did not approach zero. Given the puzzling variability in click rates, the click-rate data alone from 12 exercises over the 3.5-year period were insufficient to characterize the training effect. Therefore, in early 2016, human factors researchers employed by the institute were asked to help investigate and explain the variability in click-rate data from the trial deployment. A review of click rates and scenario types did not yield a satisfactory explanation, only more questions. Quantitative click-rate data only told us *how many* people clicked, but not *why* they clicked. We realized we needed click-decision information directly from participants to help explain why they were clicking and not clicking in their own words. Garfinkel and Lipford expressed it this way, "In order to train users to avoid phishing attacks it is necessary to first understand why people are falling for them" [9].

III. Method

To better interpret the institution's prior 3.5 years of operational click-rate data—and understand click rates for future planned exercises—we decided on a multiple methods assessment approach [24]. We knew that numbers alone did not tell the whole story and that purely quantitative click-rate data could not answer our assessment question: *Why are email users clicking or not clicking on phishing links and attachments?* Further, we needed to understand why people were or were not clicking to inform the organization's trial deployment question: *Why are click rates so variable?* We decided to use a survey instrument to probe click decisions by obtaining participant feedback following each of the next three training exercises in 2016, represented in Fig. 1 as triangle-shaped data points. The survey method allowed for immediate administration of the instrument following a click decision. Additionally, an anonymous, online survey was preferable in the workplace to provide the freedom for more honest responses compared to in-person methods. We followed the appropriate human subjects approval process for our institution.

A. Approach

We developed, tested, and fielded a web-based survey instrument for three new phishing awareness training exercises in 2016, while meeting regularly with our ITSO partner. Our multiple methods assessment used a survey instrument having open- and closed-ended responses to gather qualitative and quantitative data to complement our click-rate data. Our coding and analysis approach probed similarities and differences in survey response data between and among clickers and non-clickers.

When developing and conducting phishing exercises during 2016, the ITSO followed the same protocol as previously described for the 2012 to 2015 exercises, with the following exception: the ITSO sent a survey invitation email to each employee after they clicked or did not click the link or attachment in the phishing training email. The phishing training software tracked whether participants clicked or not. For those who clicked, the ITSO immediately sent a survey invitation email. For those who did not click, the ITSO waited a week before sending the survey invitation email (to allow them sufficient time to have seen the phishing exercise email). The survey invitation email included the names and contact information of the researchers conducting the assessment, and the appropriate link depending on the employee's click decision. Although operational security data on click rates were identifiable, survey data were confidential. Intentionally, there was no linkage between individuals and survey responses to help preserve privacy in the workplace.

B. Environment and participants

It is always important to understand the environment in which a study is conducted, but this is especially critical for a long-term operational assessment conducted in situ. The institute where this assessment took place is highly security-conscious, with yearly mandatory IT security awareness training for all staff. The OU that participated in the deployment benefited from an involved ITSO. For instance, the ITSO proactively sent emails advising people of current security scams, developed phishing awareness training that was tailored

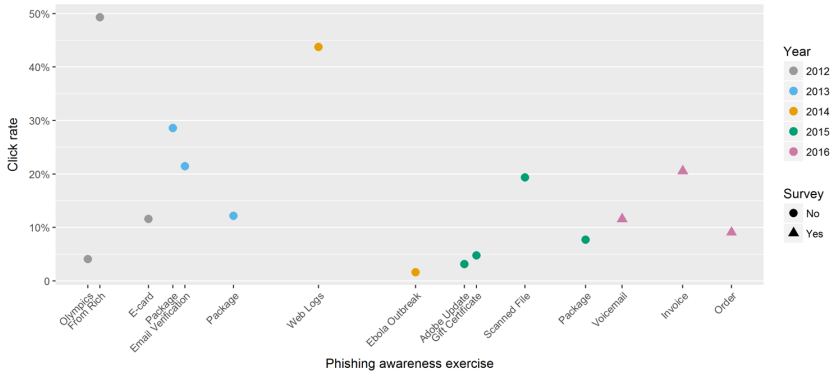


Fig 1: Phishing Awareness Exercise Click Rates, 2012 through 2016

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

specifically to the OU, reminded staff to forward suspicious emails to the security team, and generally encouraged people to be actively engaged in security. Our ITSO partner was well-known within the OU and staff knew of the phishing emails. The fact that staff were aware they were participating in a long-term phishing awareness effort differed from the studies reported in [4], [8], and likely [17], where those assessed did not know they would be phished by their respective organizations.

Participants in our assessment sample were staff at a U.S. government research institution, specifically those individuals comprising the OU previously described. The OU had good variability in age, education, supervisory experience, and time spent at the research institution. The OU was 67 % male and 33 % female, with 23 % of individuals in supervisory positions and 77 % in non-supervisory positions. Ages ranged from 25 to 66 years, *Mean* = 48.7, *SD* = 9.9. Education ranged from high school to doctorate degrees. Time working at the research institution ranged from less than one year to 39 years. The OU demographics were similar to those of the larger institution, which has approximately 3000 staff. *N*, the number of people who participated in a given phishing exercise, ranged from 66 to 73 across the three training exercises, as shown in Table I of the Results section.

C. Instrument development

In consultation with our ITSO partner, we developed an online survey instrument for the March 2016 phishing awareness training exercise. We solicited feedback on our survey from three domain experts; these were Human-Computer Interaction (HCI) researchers with significant usable security domain expertise. Based on the resulting survey data from the March exercise, and to customize the survey for upcoming exercises, we made minor refinements to the survey instrument template for use in the August and December exercises. Slight refinements were made to refer to an attachment versus a link (as appropriate given the nature of each exercise) and to include the screen image capture of the phishing email in question. We also added specific close-ended questions to further probe findings from open-ended responses in the first survey. E.g., we added a four-point Likert scale question on confidence in institutional computer security measures (with an open-ended follow-up question), and a yes/no/not sure question on perceived behavior change (also with an open-ended follow-up). To further validate the refined survey instrument [13], the instrument was reviewed by two survey experts and two additional domain experts. We also performed two formal cognitive walkthroughs with pseudo-participants (persons representative of the participant sample, but who were not part of the sample).

Importantly, the structure and content of all three surveys was consistent and nearly identical, with the aforementioned customizations made to the survey template based on the nature of the phishing email (link versus attachment). This is especially critical when considering the open-ended qualitative data from participants' impressions of the email and click decision explanations, which form the bulk of our analyses and findings. These were consistent across exercises. In order to better elicit participants' initial impressions without bias, these open-ended questions started the survey template, with close-ended questions following thereafter. Clickers received a survey version that said, "did click" and non-clickers received a version that said, "did not click" when referring to the open-ended questions about initial impressions and click/non-click decisions. That was the only wording difference between clickers and non-clickers; all other questions were identical.

Surveys were conducted via commercially-available web-based software. The survey template consisted of a mixture of open-response (3 questions) and close-ended questions (17 questions, 7 of which had open-ended follow-ups), for a total of 20 questions and an estimated response time of only 9 minutes. We felt it was important to have a short survey since participants were in the work setting and we wished to minimize time taken away from their regular tasks. Survey questions addressed initial impressions of the phishing training email, click decision explanations, device used, concern over possible consequences for clicking/not clicking, hurrying, curiosity, suspicion, confidence in the institution's computer security measures, number of emails sent and received daily, demographics, impressions of phishing awareness training and behavior change, and a final question asking for any additional comments on the survey or phishing exercises. It is important to reiterate that questions on initial impressions and click decision explanations were open-ended and consistent across exercises. Additionally, as should be clear in the Results section, there were open-ended follow-ups to questions on confidence in institutional computer security measures, consideration of consequences, and behavior change. We encourage other researchers to use our survey instrument template (Appendix Table 1) for future operational phishing research.

D. Training exercise details

Here we describe the three phishing awareness training exercises with corresponding surveys conducted during 2016. Each exercise used a different premise based on a real-world phishing threat current at that time. Two exercises were link attack scenarios, designed to mimic general malware distribution attacks. One exercise was an attachment attack scenario designed specifically to mimic the Locky ransomware attack. Each email contained cues, such as misspellings or a missing salutation, that a discerning recipient could have used to correctly identify the email as a phish. None of the three emails asked for personal information such as usernames or passwords, often a suspicion trigger [7], as such traditional phishing attacks were not the most prevalent threats at that time.

1) First exercise: new voicemail (March)

In the first exercise, the phishing training email appeared to be a system-generated message from the fictitious CorpVM <mailto:corpvm@webaccess-alert.com>. The subject line was "You have a new voicemail." The greeting was personalized with the recipient's first and last name. The body referred to an undelivered voicemail with these instructions: "To listen to this message, please click here. You must have speakers enabled to listen to the message." The body also included a reference number for the message, length of transmission, receiving machine ID, thank-you, and smaller footer text that said, "This is a system-generated message from a send-only address. Please do not reply to this email."

2) Second exercise: unpaid invoice (August)

In the second exercise, the phishing training email appeared to be from a fictitious federal employee of the same institution

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

as the recipients: Preston, Jill (Fed) <jill.preston@nist.gov>. The subject line was "Unpaid invoice #4806." The greeting was personalized with the recipient's first and last name. The body of the email said, "Please see the attached invoice (.doc) and remit payment according to the terms listed at the bottom of the invoice. Let us know if you have any questions. We greatly appreciate your prompt attention to this matter!" The email closed with the name of the fictitious federal employee ("Jill Preston"), but without any additional contact information. The attachment was labeled as a .zip, although the email body text referenced the filename "invoice_S-37644806.doc"–an extension mismatch.

3) Third exercise: order confirmation (December)

In the third exercise, the phishing training email appeared to be from the fictitious "Order Confirmation auto-confirm@discontcomputers.com" [sic] with this subject line: "Your order has been processed." There was no greeting, either personalized or generic. The body said, "Thank you for ordering with us. Your order has been processed. We'll send a confirmation e-mail when your item ships." The body also included an image of holiday packages–this email was sent in December before the holidays, order details, an order number, estimated delivery date, subtotal, tax, and order total. There was a "Manage order" button that users could click and closing text that said, "Thank you for your order. We hope you return soon for more amazing deals."

E. Data analysis

Throughout the data analysis process, we focused on answering our operational assessment question: *Why are email users clicking or not clicking on phishing links and attachments?* With our multiple methods approach, we used qualitative and quantitative survey data from participants, quantitative click-rate data captured automatically by the phishing training software, and observations by the phishing awareness training conductor—our ITSO partner.

As our open-ended survey questions were few, short, and bounded, and the response space was small, we chose to use a single coder rather than engage in group coding. Using an iterative analysis process [19], the same team member coded all open-ended survey data then reported back to the other research team members for analytic group discussions. More specifically, the coder first read through the survey data multiple times to become familiar with the content before beginning the process of coding, then prepared spreadsheets to help examine response similarities and differences between clickers and non-clickers for each question with respect to emerging themes. All research team members met regularly to discuss the ways in which codes led to particular themes, and to identify relationships amongst codes.

The coder also computed descriptive statistics for the quantitative data, comprised of click-rate data and close-ended survey data. These quantitative analyses were also shared with the larger research team and examined in conjunction with the qualitative coding during analytic group discussions. The entire team examined the qualitative and quantitative analyses for each of the three phishing exercises and compared findings across exercises, focusing on similarities and differences in clickers' and non-clickers' responses among the phishing exercises. These comparisons across exercises showed good triangulation within and across clickers and non-clickers. Finally, we performed member-checking by presenting our analyses—both qualitative and quantitative—to our ITSO partner, who provided us with contextual information about each phishing exercise.

We chose to focus the following Results section on our qualitative results and themes, as these were novel findings surrounding user context over nearly a year in a true operational setting. As previously described, our open-ended survey questions asked participants in their own words to describe their initial impressions of the phishing email and to explain why they chose to click/not click on the phishing link or attachment.

IV. RESULTS

For the three 2016 phishing exercises, Table I presents click rates–the number of recipients who clicked on the link or opened the attachment—and survey response rates. The unpaid invoice scenario had by far the highest click rate, approximately twice that of the other two scenarios. In subsequent sections, we present a summary of findings for each exercise, followed by primary themes that ran across exercises. Illustrative supporting quotes with reference codes are intermixed throughout. Each participant was assigned a reference code, denoted as ###C or ###NC for clickers and non-clickers, respectively. Numbers in the 100 series refer to the first exercise, the 200 series refers to the second exercise, and the 300 series refers to the third exercise. We focus heavily on results from the second exercise for several important reasons: 1) this phishing email mimicked the real-world Locky ransomware prevalent at that time, 2) it had the highest click rate of the three exercises, and 3) we therefore have the most survey data from clickers for that exercise.

TABLE I. PHISHING EXERCISE STATISTICS

Exercise	*n*	Phishing click rate	Survey response rate, clickers	Survey response rate, non-clickers
New voicemail	69	11.6 % (8/69)	100 % (8/8)	21.3 % (13/61)
Unpaid invoice	73	20.5 % (15/73)	66.7 % (10/15)	25.9 % (15/58)
Order confirmation	66	9.1 % (6/66)	66.7 % (4/6)	50 % (30/60)

A. Phishing premises and cues

For both clickers and non-clickers, there seemed to be an accumulation of cues that contributed to their click/non-click decisions in each exercise, with an individual's work context being the lens through which all other cues were viewed and interpreted. By user context, we mean the unique combination of an individual's work setting, responsibilities, tasks, and recent real-world events in their life: how a person experiences reality in the workplace. Most importantly, the alignment between a user's work context and the general premise of a phishing email determined whether they found the email initially believable or suspicious, which in turn influenced the cues they attended to.

1) First exercise: new voicemail (March)

Clickers perceived the email as looking legitimate and professional, and aligned with external events, such as having

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

missed a call earlier. They were unsuspicious because they were not asked for personal information, and they found the email believable because unfamiliar emails at work are common for them. For example, when explaining their click decisions, respondent 103C stated, "My phone rang earlier but I wasn't able to pick up in time," and 106C noted, "Recent talks about changing to a VOIP system." (VOIP stands for Voice over Internet Protocol.) 101C said, "The email looks professional," and 105C explained, "It looked legitimate." 104C elaborated, "The unfamiliar email is common at work, and generally not a problem. Did not trigger anything in my brain that would indicate that it was harmful. Did not ask me to give personal info like social security number etc."

In contrast, non-clickers were suspicious of the email's appearance, often describing it as "strange," "spammy," and "unusual." They also reported suspicion-raising cues in the *from* address or company name; misalignment with expectations or external events (no voicemail indicator on phone); and the unfamiliar message, never having received an email about a voicemail before. For example, 102NC said, "this had a very 'spammy' feel to it, especially since it was a service I've never heard of." 105NC noted, "it did not look like an official government email." 109NC explained, "It was from a '.com' email address instead of '.gov'." 108NC explained, "I didn't recognize the CorpVM email. I didn't have a vociemail [sic] on my phone." 105NC described, "I did not recognize the sender or the company, my voicemail is not connected to the web as far as I know." 106NC elaborated, "I am not familiar with the company, it was from an unknown caller, I hovered over the hyperlink and the url looked odd and the voicemail light was not lit up on my phone. All very suspect!" 111NC explained, "I didn't get a phone call prior to receiving the email. Made the email about voicemail very suspect." 113NC reasoned, "I've never seen a message telling me I had a new voice mail," and 101NC said, "Did not look like the normal email that I receive when I get a voice mail message."

2) Second exercise: unpaid invoice (August)

As in the first exercise, responses from non-clickers differed greatly from those of clickers. For most respondents whose work responsibilities included handling financial matters—such as paying vendors or dealing with contracts and grants—the email was particularly believable and concerning given their individual user context. This was evident in clickers' comments on their initial impressions of the email and their decisions to open the attachment. For example, 202C reported thinking, "That I had an unpaid invoice from when I assisted my division with credit card orders," and 203C said, "that I missed a payment on one of my contracts." 209C was concerned, "I had forgotten about an invoice for either a contract or purchase order." 207C questioned, "What did I order that I haven't paid for?" and 210C noted, "I pay invoices so I was wondering what invoice this was that did not get paid."

Clickers also referred to recent real-world events that were congruous with the unpaid invoice email. 204C explained, "I had just talked to someone in accounts payable the previous day who told me they were new. I thought they were reaching out by email this time bc I saw the .gov" and 208C said, "We recently had a legit email from a vendor regarding unpaid invoices so I thought this may be another one. I work with private vendors so it looked legit."

The fictitious sender's @nist.gov email address was a particularly believable cue for clickers, with comments like "it came from a Fed" (203C) and "The email was from an internal email address." (207C) Similarly, 210C reasoned, "From a NIST employee, figured she worked in AR and/or finance." (AR stands for Accounts Receivable)

In general, non-clickers found the premise of this email particularly suspicious given their user context: either their work responsibilities did not include finances, or if they did, they found enough suspicious cues to prevent them from opening the attachment. For non-clickers, the mismatches with user context and expectation were particularly salient in this exercise, triggering enough suspicion to prevent them from opening the attachment. 203NC said, "My first impression was to ask why I was getting an email regarding payment of an invoice because I don't have a credit card or purchasing authority. It made me suspicious." 204NC elaborated, "I immediately thought this seemed like a strange email to receive. I don't usually get emails about invoices that need to be paid, and certainly not from a NIST email address. I also thought it was strange that the attachment was a zip file." 206NC noted, "I definitely thought it was suspicious since I don't get any invoice related stuff." 207NC said, "I don't know this person and if they are from NIST they should not be sending an invoice this way." 212NC explained, "I don't deal with invoices or anything having to do with accounts payable or accounts receivable and haven't recently purchased anything." 213NC said, "The email did not match something that a federal employee would write. It seemed more like an email from a vendor."

Importantly, although some non-clickers had finance-related responsibilities, they still found enough suspicious cues that they chose not to open the attachment. For instance, 202NC said, "The format of the email did not match the format of other invoices sent by accounts payable. The invoice number did not match the series used in either of the projects that I manage. The email was also addressed to my last name (not the usual format). The attachment was a zip file. Invoices are never in zip files."

Like clickers, non-clickers also noticed and referred to the .gov email address cue. However, they performed additional fact-checking that clickers did not: they searched for the fictitious Jill Preston in the employee directory. 204NC said, "I searched for the sender in the directory and could not find them." 205NC explained, "I looked up Jill Preston in the user directory and the person didn't exist." 206NC said, "I looked up the name in [the email client]." 207NC noted, "Looked the person up in the directory and did not find them." 213NC noted, "I searched for the person in the phone directory and did not find anyone."

In a quote that perfectly illustrates a non-clicker's progression from general suspicion, to noticing a specific cue, to engaging in fact checking, and finally to reporting, 201NC explained, "…upon re-reading the email I became very suspicious. The email references a .doc attachment, but the attachment was a .zip file. After noticing that, I checked the NIST directory and saw that there was not a Jill Preston (Fed) at NIST. I immediately forwarded to my ITSO."

3) Third exercise: order confirmation (December)

Although this exercise had the lowest click rate of the three, the importance of a user's individual work context was again clear. When the premise of the email seemed believable given

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

user context, respondents clicked on the link. This match between an individual's reality and the email was clear for clickers. 302C explained, "We have some items on back order so I thought that this may be one of those so I clicked on it to see what the item was." 303C said, "I have several orders open and sometimes get email shipping confirmations," and "I did get a legitimate one that day as well."

In contrast, the premise of the email did not match reality for non-clickers' work contexts. For instance, 319NC explained, "This is weird because I don't place orders or pay for them." 305NC said, "This looks like a personal order. I never use my work email for ordering personal items." 310NC noted, "Suspicious since I didn't remember ordering anything." 330NC said, "Became cautionary after realized that this did not fit anything I had recently purchased."

As in the preceding two exercises, non-clickers focused more on suspicion-raising cues and reasoning through the cues and email content when making their decisions not to click the link. 301NC explained, "After recognizing an amazon imitation, I looked at the email address 'discountcomputers.com'. Never heard of it, or bought anything from there." 310NC also referred to the email address: "Discount was mispelled [sic] on the sender's email address." Similarly, 304NC said, "I noticed that 'discount' was spelled incorrectly in the email address."

Non-clickers also referred to the lack of specificity in the email's content. For example, 320NC explained, "As an unknown subject from an unknown source, with rather vague & generic content, it was obviously not legit." 309NC said, "There was no information identifying the company that was sending the email and there was no product information." 305NC noted, "This does not look like it comes from our Ordering systems. It had no descriptive organization or company."

As this exercise took place in December, the overall holiday shopping context was important, something that came up in comments from both clickers and non-clickers. For example, 309NC said, "I have been ordering some Christmas stuff but I have always used my home email so I was a little confused as to why this was coming to my work email."

B. Themes across exercises

1) User context

Across all three phishing exercises, a user's individual work context was key in understanding an individual's click decision. The importance of alignment/misalignment with a user's work context, expectations, and external events was clear in each exercise. Whereas clickers found the premise of the phishing emails to be particularly believable given their user context, non-clickers did not. Clickers tended to focus on compelling cues, such as the basic premise of the emails, or the .gov email address. However, non-clickers focused on suspicious cues like misspellings and unexpected attachment types. Because their user context led them to question the premise of an email, non-clickers reported performing additional fact-checking, such as searching for the sender in the employee directory.

For clickers, the emails were plausible enough given their work context that deeper thought and analysis were not triggered, indicating more surface level thinking. Even if something did seem out of the ordinary to them, it was insufficient to trigger alarm bells, or any suspicion raised was counteracted by other believable cues, such as the presence of a .gov email address. For instance, 201C said, "Seemed a little out of the ordinary but legit email address." 104C said, "thought it was odd, but didn't connect the dots between phishing scheme and voicemail." In contrast, for non-clickers, the emails were suspicious enough that deeper thought was triggered: they questioned multiple cues, reasoned carefully about the emails, and even took additional steps to check facts when possible. Non-clickers also reported re-reading emails, for instance, 201NC said, "Checked address and saw (FED) so automatically assumed it was legit. After reading email, it was a little off, so I reread a few times."

2) Strategies and fact-checking

Across exercises, non-clickers described specific strategies they used and cues they looked for. For example, 326NC said, "I check multiple items, the from, the address when you mouse over, the content in general, look for misspelled words, strangely worded emails." 101NC described, "I have certain rules that I follow, I don't click on any links when I don't know who the email is from, and I don't click on any links when the email is from someone that I know but there is no message from that person." 325NC elaborated, "It reminds me of what to look for like a peculiar email address, odd salutation, bad grammar, and a sense of urgency in the message to make you do something that you normally wouldn't do. Because of this training, I am highly suspicious and my 'Spidey Sense' tingles whenever I see one of these emails."

In addition to following strategies and behavioral rules, non-clickers also engaged in fact-checking tailored to the premise of the phishing exercise. In the first exercise, they checked the voicemail light on their phone. In the second exercise, they searched for the fictitious Jill Preston in the employee directory. In the third exercise, they thought back through whether they had recently ordered anything. There were also several unique fact-checking strategies reported. For the second exercise, 103NC explained, "I checked the serial number of my work [mobile phone] to see if it was the same number as in the email." (Presumably this person was referring to the "receiving machine's ID" listed in the exercise email). Additionally, 304NC reported a clever fact-checking strategy regarding the sales tax listed in the third exercise email: "The MD sales tax is 6 %. I calculated the tax based on the cost ($59.97), and the sales tax listed in the email is greater than 6 %."

While clickers did not report these types of fact-checking strategies in relation to the phishing emails, several engaged in fact-checking regarding the survey itself: for the second phishing exercise, three participants called the NIST researchers to verify that the survey was not a phishing attempt. When speaking with us, they said they had just been caught by the training and wanted to check this was not another phish.

3) Concern over consequences

Concern over potential consequences differed greatly between clickers and non-clickers. Across the three exercises, clickers were often concerned over consequences or repercussions arising from *not* clicking, such as failing to act or failing to be responsive. Referring to the voicemail phishing exercise, 105C explained, "I am always interested in ensuring that I get any messages and act on them. It could have been my supervisor or other person requiring an action on my part."

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

Clickers were particularly concerned over the consequences of not addressing an unpaid invoice: if they did not click to open the attachment, they could not pay the invoice nor could they forward it to the correct person for payment if it was not their direct responsibility. The unpaid invoice exercise also had the highest click rate, nearly double that of the other two exercises. The fact that an actual unpaid vendor invoice had recently been an issue in that OU no doubt contributed to clickers' increased concern over the email. 201C referred to "repercussions for not remitting payment," 203C responded, "potential to miss contract payments," 207C lamented, "I would have an unpaid invoice," and 208C worried, "I may miss a legitimate complaint/issue." 208C explained, "If true, I would be the person who would have to address the issue." 202C noted, "In the past 10 months I had 2 cancelled orders and wondered if it had gone through."

Across all three exercises, non-clickers were more concerned with consequences that could arise from clicking, such as downloading malware and viruses. For example, 103NC said, "I was concerned something might be downloaded onto my computer or I could get a virus." 106NC explained, "If this was a phishing email, I could be the person that allows someone access and they could potentially infect or steal information." 204NC said, "I considered that it might contain a virus or other malware." 210NC said, "I considered a hacker planting a listening bug onto the NIST systems." 211NC explained, "It did not look like a legit email so clicking on the link could have lead [sic] to virus, spyware, malware, etc." Similarly, 306NC said, "I thought if I clicked I could get a virus or have malware put on my computer."

4) *Confidence in institutional computer security measures*

Clickers seemed quite confident in the institution's security measures. For example, 101C stated, "I thought NIST security system can filter phishing emails." 105C said, "We are within a firewall at NIST?" 303C noted, "I do not get spam or junk emails (very often if at all), which tells me NIST is very pro-active in stopping them before we get them."

Non-clickers seemed to have a more tempered view, frequently referring to the idea that some phishing emails will always get through the filters. For example, 201NC explained, "Because of the widely varied nature of the work here, I am not sure it is possible to block out all phishing attempts." 202NC said simply, "It's impossible to catch them all." 213NC explained, "My feeling is that some phishing emails will get through no matter how good the security measures are." 320NC, "I think NIST puts serious effort into computer and internet security but things are always going to get through on occasion." 312NC noted, "There is no perfect security. The best NIST can hope for is to mitigate the number of attacks." 324NC said, "the attacks are constant, some will get through." Interestingly, 203NC referred to progress for both the institution and the hackers, "NIST has made a lot of progress in stopping phishing emails, and has trained staff well. However, I know that hacking continues to advance, too."

V. DISCUSSION

A. *Benefits of operational data*

The benefits of long-term, in situ operational phishing data are many. Specifically, the benefit of ecological validity is extremely valuable. By conducting an assessment using a sample of real-world staff working in their normal work environments, we were able to obtain a high degree of ecological validity lacking in many assessments. Importantly, our sample was varied in terms of age, education, supervisory experience, and years spent working at the institution. Additionally, we surveyed exercise participants almost immediately after they clicked, in contrast to [4], where many months elapsed between the time participants received the initial phish and were interviewed about it. We also addressed the "adversary modeling challenge" identified by Garfinkel and Lipford in [9], by using phishing emails modeled on real-world threats current with each exercise. As the nature of phishing continues to change, it is imperative to model attacks after real-world threats.

We believe that our long-term, in situ assessment with a varied sample should be meaningful and compelling for usable security researchers and security practitioners alike. Our workplace setting allowed us to both confirm findings from laboratory settings and unearth new findings that were only possible with a real-world setting.

B. *User context: an explanatory variable in phishing susceptibility*

A user's context—in this case, their individual work context—is the lens through which they interpret both the general premise of an email, as well as specific cues within the email. Through three phishing awareness training exercises and corresponding surveys, conducted in situ over a 10-month period, we found clickers and non-clickers interpreted the premise of phishing emails very differently depending on their individual work context. When a user's work context aligned with the premise of the email, they tended to find the premise believable and attended to compelling cues. In contrast, when a user's work context was misaligned with an email's premise, they tended to find the premise suspicious and they instead focused on specific suspicious cues.

Once we identified user context as an explanatory factor, we then saw that user context alignment coincided with depth of processing and differences in concern over consequences between clickers and non-clickers, which are discussed next. From respondent feedback, the initial read of the email by clickers seemed to produce a reaction that aligns with findings of Wang et al., where "Visceral triggers reduce the recipients' depth of information processing and induce recipients to make decision errors" [27] and Kumaraguru et al. in [15]. On the other hand, non-clickers reported a reaction of suspicion to the premise misalignment with their user context, helping them attend to phishing deception cues. We also discuss the operational setting and user context.

1) *Depth of processing*

Clickers and non-clickers seemed to process the phishing awareness training emails quite differently. When the general premise of a phishing email was believable for clickers, they did not then engage in deeper processing; they did not look for deception indicators, nor did they question the email's legitimacy. However, we saw evidence of extremely efficacious cue detection and utilization strategies on the part of non-clickers, such as checking for misspelled words, grammatical errors, mismatches, and so on. Our data also show that non-clickers report engaging in additional fact-checking behaviors, such as searching for the sender in the employee directory.

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

From a cognitive science perspective, this makes perfect sense: why waste effort and time on double-checking something that seems perfectly legitimate? In contrast, when the general premise of a phishing email was suspicious for non-clickers or they attended to specific suspicious cues, they engaged in deeper processing to confirm that the email was indeed a phish. For example, non-clickers reported re-reading the emails and engaged in additional fact-checking to follow-up on suspicious cues, such as checking their phone lights and searching for Jill Preston in the institution's employee directory. This difference in surface versus deep processing ties back to System 1 and System 2 thinking; with System 1 being fast, intuitive, and emotional, relying heavily on habits and heuristics, and System 2 being slower, more deliberative, and more logical [12]. This is in no way to say that clickers and non-clickers differ in depth of processing in general, only that for these particular phishing awareness exercises set in the real world, their survey responses showed evidence of differential processing.

2) *Consequence considerations*

Overall, both clickers and non-clickers provided feedback that conveyed an attitude of conscientiousness in the workplace. Despite that general conscientiousness, concern over consequences typically aligned with the click decision. Overall, clickers tended to be more concerned with potential consequences that could arise from *not* clicking: failing to act, seeming unresponsive to an email, or not addressing a legitimate issue. This was especially true for clickers in the unpaid invoice phishing premise, and seemed exacerbated by recent events where an unpaid invoice had been an issue for the OU participating in the awareness training exercises. In contrast, non-clickers were concerned with potential consequences that could arise from clicking: malware, viruses, and so on.

3) *The operational setting and user context*

It is important to realize that despite being part of the same approximately 70-person OU within the institute, staff in our assessment had very different *individual* work contexts. Responsibilities within the OU ranged from financial matters, such as processing orders, invoices, grants, contracts, and payments; to organizational safety matters relating to radiation, health and environmental issues; to program development and compliance. Even within the subset of staff members who shared somewhat similar responsibilities, individual work contexts varied. For example, within the subset of staff who had financial responsibilities, individual contexts varied depending on whether someone was recently working on a particular invoice and with whom they were working.

In addition to an individual's work context, another work context is the shared work context. For example, consider the collective awareness among our assessment participants regarding a recent unpaid vendor invoice. This shared awareness of a recent workplace issue affected participants' concern over the consequences of an unpaid invoice and may have contributed to elevated click rates for that phishing exercise. Though not uniquely a work context, an additional example of a shared context was the holiday context in the third exercise–that exercise leveraged the fact that more people place online orders near the holidays.

Such variability in individual work contexts, that at times partially overlaps with others' work contexts, is likely the case at other institutions as well. It would be rare to find an organization where every member of the staff has exactly the same, or even completely different, tasks in at least medium-sized or larger organizations. Different people will be more or less susceptible to a particular phishing email's premise based in part on their individual work environments, tasks, and responsibilities.

C. *Revisiting prior quantitative data*

Based on the primary contribution from our 10-month operational assessment—showing the importance of individual user context in explaining phishing email click decisions—we are now able to better interpret the previously puzzling variability in click rates observed across the initial 3.5 years of phishing awareness training exercises at a U.S. government institution. Although we do not have user feedback from earlier exercises, the lens of the user context and the premise of each prior exercise together give new insight in understanding click results. Fig 1. shows click-rate data. Several data points deserve special mention. The exercise labeled 'From Rich' had a high click rate (49.3 %) and used a message that was tailored to mimic an actual spear phishing attack on the organization just prior to the exercise. The premise in this exercise aligned extremely well with the communication practices of the organization and the culture of conscientiousness. The phish labeled 'Web Logs' purported to be an automated message citing a violation of the institute's internet policy; it also had a high click rate (43.8 %). The premise this time aligned with the fact that the organization had such a policy and staff were aware of it. The scenario with the lowest click rate, 1.6 %, is labeled 'Ebola Outbreak.' Its low click rate is likely due to the circumstance that the institute's spam filtering placed the training phish in people's spam folder; despite this, one person saw it there and clicked its attachment. Knowing that people in this OU were responsible for organizational health issues at the institute and noting that the timing coincided with the Ebola crisis of 2014, we better understand the premise alignment. Click rates from the other exercises during 2012 to 2015 are between these extremes; each had plausible premise alignment with some staff members' job responsibilities within the OU or an external event that was relevant to the individual. This provides a plausible explanation why the institute's ITSND observed such intriguing variability in click rates across the initial 3.5 years of their trial deployment.

D. *Findings with actionable operational implications*

With data from an operational setting, we have the rare opportunity to make novel observations that have operational implications as well as confirm findings from other studies that hold in a new operational environment. We report on three findings that are immediately actionable by security practitioners.

1) *Setting realistic expectations of institutional security*

We found that participants in our assessment, especially clickers, expressed a great deal of confidence in the institution's technological security measures, perhaps too much confidence. While it is good that staff are aware there are indeed institutional security measures in place, having too much faith in such mechanisms can be dangerous if it leads staff to a false sense of total security. If people are overconfident in, or overly reliant on, institutional security measures, this may have an undesirable

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

effect on their clicking behaviors. They may not be concerned about clicking if they believe they are already fully protected. Organizations should consider explicitly educating their employees that no technological solution is completely infallible, especially reactive ones like those still commonly in use. Our operationally-situated findings of staff confidence in the institution's security measures provide confirmation to those described in [6] and [7], both of which were performed in other settings.

2) *The changing nature of phishing attacks*

Our exploration of click decisions revealed that some members of the sample did not know that the nature of phishing has evolved beyond the traditional requests for sensitive information such as usernames and passwords. Others, such as [7], have reported that users who know the definition of phishing are statistically less likely to fall for phishing than those who do not know the definition. This highlights the operational opportunity to ensure staff remain aware of how phishing attacks continue to evolve.

3) *Gamification*

Our ITSO partner observed emergent competition or *gamification* of the phishing awareness training exercises over the years, where people would try to beat their colleagues and be the first person to spot the training phishing email. In an effort to have each member in the OU make their training click decisions without co-worker aid, the ITSO cautioned people not to warn others of the emails. However, friendly competition among colleagues may be an unintended benefit to conducting long-term phishing awareness training exercises. Since phishing messages missed by technological solutions are often found by individuals reporting suspected phish or after negative consequences arise, an increase in the number of those reporting phishing messages, especially early reporting, should provide meaningful benefits to individuals and the information technology security incident-response teams tasked with detecting, containing, eradicating, and recovering from infected machines and networks [5]. Organizations have the opportunity to use that friendly competition to encourage reporting suspected phishing messages.

E. *Implications for predicting phishing susceptibility*

Our data establish user context as an explanatory variable in phishing susceptibility. Given the long-term, in situ nature of our data across a wide range of ages, education levels, and work experience, we believe our data are compelling and comprehensive. Of course, all new findings should be validated by additional studies, as should this one. However, studies that attempt to examine user context must be set in participants' real-world settings; laboratory settings cannot provide a sufficient level of ecological validity for this particular phenomenon. Additionally, models that attempt to predict phishing susceptibility such as those documented in [23] and [26] should further explore the user work context identified in this paper. This is in addition to other phishing susceptibility factors such as personality and individual risk tolerance, which others have studied, for example [3], [7], [10], [14].

VI. Conclusions

The problem of phishing is not yet solved and the impact on individuals and organizations continues to grow as attacks become increasingly sophisticated. We have uncovered another piece of the phishing puzzle through an examination of long-term, operationally-situated data captured during embedded phishing awareness training exercises conducted at a U.S. government institution. In this paper, we focus on three exercises conducted in 2016, each having participant feedback. For all 15 exercises in the 4.5-year span (2012 through 2016), the phishing training emails modeled real-world phishing campaigns, and participating staff were in their normal work environments with their regular work loads, providing ecological validity.

This data has revealed the novel finding that the alignment of user context and the phishing message premise is a significant explanatory factor in phishing susceptibility, impacting both an individual's depth of processing and concern over consequences. Additionally, our data provide an operationally-relevant perspective from users regarding other factors affecting their click decisions. We believe that the rare opportunity to collect data surrounding phishing click decisions in the workplace coupled with ecologically valid conditions over time provided the crucial elements for these findings to surface. Because of the scarcity of operationally-situated data, we also report on findings we believe to be actionable now in operational settings. Our long-term operational data offer significant contributions to the existing corpus of phishing literature with implications for both future research and operational security.

Acknowledgements

The authors gratefully acknowledge the institution's ITSND and our ITSO partner for their support throughout this project, and thank anonymous reviewers for their comments.

Disclaimer

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

References

[1] G. Aaron and R. Rasmussen, "Global phishing survey 2016: Trends and domain name use," Anti-Phishing Working Group, June 2017. https://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf (Accessed Aug 2017).

[2] A. Almonnani, B.B. Gupta, Samer Atawnech, A. Meulenberg, and E. Almomani, "A survey of phishing email filtering techniques," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, Fourth Quarter 2013.

[3] M. Blythe, H. Petrie, and J.A. Clark, "F for fake: Four studies on how we fall for phish," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM. May 2011, pp. 3469-3478.

[4] D. Caputo, S.L. Pfleeger, J. Freeman, and M. Johnson, "Going spear phishing: Exploring embedded training and awareness," in IEEE Security & Privacy, 12(1), January 2014, pp. 28-38.

[5] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, "Computer security incident handling guide," NIST Special Publication 800-61, rev. 2, 2012, http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf (Accessed Oct 2017).

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

[6] D. Conway, R. Taib, M. Harris, K. Yu, S. Berkovshky, and F. Chen, "A qualitative investigation of bank employee experiences of information security and phishing," in Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), USENIX Association, 2017, pp. 115-129.

[7] J.S. Downs, M. Holbrook, and L.F. Cranor, "Decision strategies and susceptibility to phishing," in Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS '06), ACM, pp. 79-90.

[8] A.J. Ferguson, "Fostering e-mail security awareness: The West Point carronade," EDUCASE Quarterly, 2005, 1, pp. 54–57.

[9] S. Garfinkel and H.R. Lipford, Usable security: History, themes, and challenges. Synthesis Lectures on Information Security, Privacy, and Trust, E. Bertino and R. Sandhu, Eds., 5(2), Morgan and Claypool Publishers, 2014, pp. 4-6, 55-65.

[10] T. Halevi, N. Memon, O. Nov, "Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks", January 2, 2015, https://ssrn.com/abstract=2544742 or http://dx.doi.org/10.2139/ssrn.2544742 (Accessed Oct 2017).

[11] J. Hong, "The state of phishing attacks," Communications of the ACM, 55:1, January 2012, pp. 74-81.

[12] D. Kahneman, Thinking, Fast and Slow. New York: Farrar, Straus and Giroux, 2011.

[13] K.L.K. Koskey, T.A. Sondergeld, V.C. Stewart, and K.J. Pugh, "An applications of the mixed methods instrument development and construct validation process: Transformative experience questionnaire," Journal of Mixed Methods Research, 2016, 1558689816633310.

[14] P. Kumaraguru, Y. Rhee, S. Sheng, S Hasan, A. Acquisti, L.F. Cranor, and J. Hong, "Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer," in Proceedings of the anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit, (eCrime '07), ACM, October 2007, pp. 70-81.

[15] P. Kumaraguru, S. Sheng, A. Acquisti, L.F. Cranor, and J. Hong, "Lessons from a real world evaluation of anti-phishing training," in e-Crime Researchers Summit, IEEE, October 2008, pp. 1-12.

[16] P. Kumaraguru, J. Cranshaw, A. Acquisti, L.F. Cranor, J. Hong, M. Blair, and T. Pham, "School of phish: A real-world evaluation of anti-phishing training," in Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09), ACM, July 2009, article 3.

[17] New York State Office of Cyber Security & Critical Infrastructure Coordination, "Gone phishing... a briefing on the anti-phishing exercise initiative for new york state government: Aggregate exercise results for public release," 2005. (Not available online. Additionally, the NYS Office of Information Technology Services no longer has the report–contacted 8/28/2017).

[18] J. Nicholson, L. Coventry, and P. Briggs. "Can we fight social engineering attacks by social means? Assessing social salience as a means to improve phish detection," in Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), USENIX Association, 2017, pp. 285-298.

[19] J. Saldaña, The Coding Manual for Qualitative Researchers, 2nd ed., Sage Publications, 2013.

[20] SonicWall, "2017 SonicWall annual threat report," 2017. https://www.sonicwall.com/docs/2017-sonicwall-annual-threat-report-white-paper-24934.pdf (Accessed Aug 2017).

[21] Stanford University, "University IT launches phishing awareness service," 2016. https://uit.stanford.edu/newsletter/university-it-launches-phishing-awareness-service (Accessed Aug 2017).

[22] Symantec, "Internet security threat report," vol. 22, April 2017. https://www.symantec.com/security-center/threat-report (Accessed Aug, 2017).

[23] F.P. Tamborello, K.K. Greene. "Exploratory Lens Model of Decision-Making in a Potential Phishing Attack Scenario." National Institute of Standards and Technology Interagency Report, NISTIR 8194, October 2017, DOI: https://doi.org/10.6028/NIST.IR.8194

[24] C. Teddlie and A. Tashakkori, Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches In The Social and Behavioral Sciences. Thousand Oaks, CA: Sage, 2009.

[25] Trend Micro, "Spear phishing 101: What is spear phishing?," September 2015. https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/spear-phishing-101-what-is-spear-phishing Accessed Aug 2017).

[26] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H.R. Rao. "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model." *Decision Support Systems*, vol. 51 no. 3, March 2011, pp. 576-586.

[27] J. Wang, T. Herath, R. Chen, A. Vishwanath, and H.R. Rao, "Phishing susceptibility: An investigation into the processing of a targeted spear phishing email," in IEEE Transactions on Professional Communication, vol. 55, no. 4, December 2012, pp. 345-362.

APPENDIX

This appendix contains a detailed description of the survey instrument template, including customizations, questions, response options, and show/hide question logic for follow-up questions. Survey questions are shown in Table I below. Gray highlighting is used to indicate page breaks in the survey.

A. Landing page

Each survey started with a landing page that briefly described the nature of the survey, the estimated time it would take a respondent to complete the survey, and the potential risks and benefits associated with taking the survey. The landing page also contained the researchers' contact information.

B. Customizations

As described in the Method section, the survey template was created such that it could be easily customized based on the nature of the phishing exercise. After the landing page, the first page of the survey was customized by adding a screenshot of the phishing email at the top. Open-ended questions about initial impressions (Q1) and click decisions (Q2) were together on the first page. Q2 was customized depending on whether the phishing email contained a link or an attachment, and depending on whether a respondent was a clicker or a non-clicker. Q2 customizations were as follows:

[Clickers, link] What made you decide to click the link?

[Non-clickers, link] What made you decide not to click the link?

[Clickers, attachment] What made you decide to open the attachment?

[Non-clickers, attachment] What made you decide not to open the attachment?

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

TABLE I. SURVEY QUESTIONS.

	Primary question			Follow-up question			
Q#	***Question***	***Response type***	***Response options***	***Show/hide logic***	***Question***	***Response type***	***Response options***
1	*Think back to when you first saw the email above in your inbox. What were your initial impressions?*	Open-ended		n/a			
2	*What made you decide not to click the link?* Note that Q2 was customized for link/attachment and clicker/non-clicker (as described above).	Open-ended		n/a			
3	*What type of device were you using when you opened the email?*	Radio buttons	Desktop computer Laptop Tablet Smartphone Don't remember Other – Write In: _____	n/a			
4	*Were you in a hurry when you opened the email?*	Radio buttons	No Yes Don't remember	n/a			
5	*How thoroughly, if at all, did you read the email?*	Radio buttons	Didn't read it Read some of it Read it quickly Read it thoroughly Don't remember	n/a			
6	*What were you doing around the time you opened the email?*	Radio buttons	Focused solely on checking email Checking email while primarily focused on another task Don't remember Other – Write In: _____	n/a			
7	*Did anything in the email seem pertinent to you?*	Radio buttons	No Yes Not sure if pertinent Don't remember				
				Shown only if "yes" or "not sure"	*In what way?*	Open-ended	
8	*Did you consider any possible consequences for clicking on the email link?*	Radio buttons	No Yes Not sure				
				Shown only if "yes"	*What possible consequences did you consider?*	Open-ended	
				Shown only if "yes"	*How concerned were you about these consequences you listed?*	Radio buttons	Not at all concerned Somewhat concerned Very concerned Don't remember if I was concerned at the time
9	*Did you consider any possible consequences for <u>not</u> clicking on the email link?*	Radio buttons	No Yes Not sure				
				Shown only if "yes"	*What possible*	Open-ended	

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

	Primary question			Follow-up question			
Q#	***Question***	***Response type***	***Response options***	***Show/hide logic***	***Question***	***Response type***	***Response options***
					consequences did you consider?		
				Shown only if "yes"	*How concerned were you about these consequences you listed?*	Radio buttons	Not at all concerned Somewhat concerned Very concerned Don't remember if I was concerned at the time
10	*How curious, if at all, were you about the email?*	Radio buttons	Not at all curious Somewhat curious Very curious Don't remember				
				Shown for all responses but "don't remember"	*Why or why not?*	Open-ended	
11	*When thinking about the email, how suspicious, if at all, were you?*	Radio buttons	Not at all suspicious Somewhat suspicious Very suspicious Don't remember				
				Shown for all responses but "don't remember"	*Why or why not?*	Open-ended	
12	*How confident are you in NIST's computer security measures to prevent phishing emails from reaching you?*	Radio buttons	Not at all confident Somewhat confident Confident Very confident				
				Shown for all responses	*Explain.*	Open-ended	
13	*Overall, do you feel like the phishing awareness training emails have changed your email behavior?*	Radio buttons	No Yes Not sure				
				Shown only if "yes"	*How has the phishing training changed your behavior?*	Open-ended	
				Shown only if "no" or "not sure"	*Why?*		
14	*About how many emails do you receive in a typical work day?*	Radio buttons	0 1-5 6-10 11-20 21-30 31-40 41-50 51 or more Prefer not to answer				
15	*About how many emails do you send in a typical work day?*	Radio buttons	0 1-5 6-10 11-20 21-30 31-40 41-50 51 or more Prefer not to answer				

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

	Primary question			Follow-up question			
Q#	***Question***	***Response type***	***Response options***	***Show/hide logic***	***Question***	***Response type***	***Response options***
16	*What is your gender?*	Radio buttons	Female Male Other identification Prefer not to answer				
17	*What year were you born?*	Drop-down	Options ranging from 1931 to 1998 Prefer not to answer				
18	*What is your highest completed level of education?*	Radio buttons	Highschool or GED Associate's Degree Bachelor's Degree Master's Degree Doctorate Degree MD Other – Write In: _____ Prefer not to answer				
19	*How long have you been working at NIST?*	Drop-down	Options ranging from Less than 1 year to 50 years More than 50 years Prefer not to answer				
20	*Any additional comments on this survey or the phishing exercises?*	Open-ended					

Greene, Kristen; Steves, Michelle; Theofanos, Mary; Kostick, Jennifer.
"User Context: An Explanatory Variable in Phishing Susceptibility."
Paper presented at Workshop on Usable Security (USEC) at the Network and Distributed Systems Security (NDSS) Symposium 2018, San Diego, CA, United States. February 18, 2018 - February 21, 2018.

Compositional Imaging and Analysis of Late Iron Age Glass from the Broborg Vitrified Hillfort, Sweden

Edward P. Vicenzi[1,3], Carolyn I. Pearce[2], Jamie L. Weaver[3], John S. McCloy[4], Scott Wight[3], Thomas Lam[1], Scott Whittaker[5], Rolf Sjöblom[6], David K. Peeler[2], Michael J. Schweiger[2], and Albert A. Kruger[7]

[1.] Museum Conservation Institute, Smithsonian Institution, Suitland, MD, USA
[2.] Pacific Northwest National Laboratory, Richland, WA, USA
[3.] National Institute of Standards and Technology, Gaithersburg, MD USA
[4.] School of Mechanical and Materials Engineering, Washington State University, Pullman, WA USA
[5.] National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
[6.] Luleå University of Technology, Luleå, Sweden
[7.] Department of Energy, Office of River Protection, Richland, WA USA

Numerous Iron Age hillforts were constructed throughout Europe on high ground to serve ancient settlements [1]. The edifice walls of a small fraction of hillforts were vitrified as a result of high temperature activity, [2]. Swedish hillfort glasses from the Broborg Site near Uppsala, Sweden have recently been proposed as an analogue material to inform long term nuclear waste storage (Fig. 1A) [3]. As part of that effort, a fragment of the Broborg hillfort wall was embedded and polished prior to examination by x-ray methodologies to determine its composition and microstructure (Fig. 1B).

A Bruker M4 Tornado with dual Bruker XFlash 6|60 detectors was used to collect hyperspectral micro-XRF data with a Rh source and a 20 um (Mo K_α) polycapillary optic at two energies: 1) 25 keV/no filter, and 2) 40 keV with a 12.5 μm Al source filter [4]. An FEI Apreo scanning electron microscope (SEM) was used to collect an electron image montage spanning the specimen's polished surface, and electron beam-excited hyperspectral x-ray data was collected using dual Bruker XFlash 6|60 detectors at 15 keV.

Object-scale major element and electron imagery reveals the specimen is comprised of 3 principal chemical phases, an Fe-rich glass (dark), an Fe-poor glass rich in alkalis (milky), and quartz (Fig. 2A,E). Trace element imaging depicts unreacted zircon throughout the sample, Zr and V enrichment in the Fe-rich glass (Fig. 2B-C), and Rb enrichment in the Fe-poor glass (Fig. 2D). A more detailed view of the boundary between the 2 glass zones shows heavily embayed quartz, residual CaAl-silicate zones within the Fe-poor glass, FeMgAl spinel crystallization, and quench-crystallization in the Fe-rich glass (Fig, 3). Major element compositions of the glasses are broadly consistent with previous studies (Table 1) [4,5]. Importantly, the new trace element compositions of the glasses provide an opportunity to test a linkage with the bulk chemistry of lithologies found at Broborg to better understand the ancient melting processes via an elemental ratio comparison.

References:
[1] J McIntosh, Handbook to life in Prehistoric Europe. (Oxford University Press, 2009)
[2] FB Wadsworth et al. J Archaeol. Sci. **67** (2016), p. 7-13
[3] R Sjöblom, H Ecke, & E Brännvall, Intl. J. Sust. Dev. Planning (2013).
[4] Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST or DOE.
[5] P Kresten & B Ambrosiani J. Swedish Antiquarian Res. **87** (1992)
[6] JL Weaver et al. Intl. J. Appl. Glass Sci. (*in review*)

Weaver, Jamie; Wight, Scott; Pearce, Carolyn; McCloy, John; Lam, Thomas; Sjoblom, Rolf; Peeler, David; Schweiger, Michael; Kruger, Albert; Whittaker, Scott; Vicenzi, Edward.
"Compositional Imaging and Analysis of Late Iron Age Glass from the Broborg Vitrified Hillfort, Sweden."
Paper presented at Microscopy & Microanalysis 2018, Baltimore, MD, United States. August 5, 2018 - August 9, 2018.

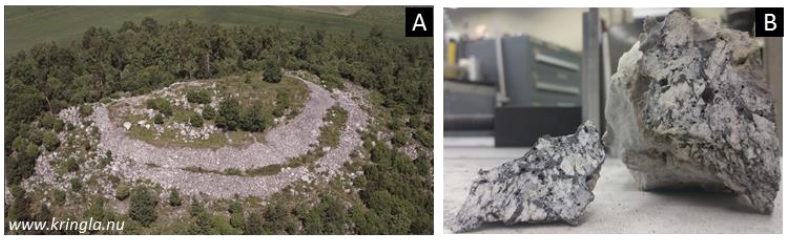


Figure 1. A) Oblique aerial view of the Broborg hillfort. B) Sectioned hand specimen image of specimen from vitrified wall.

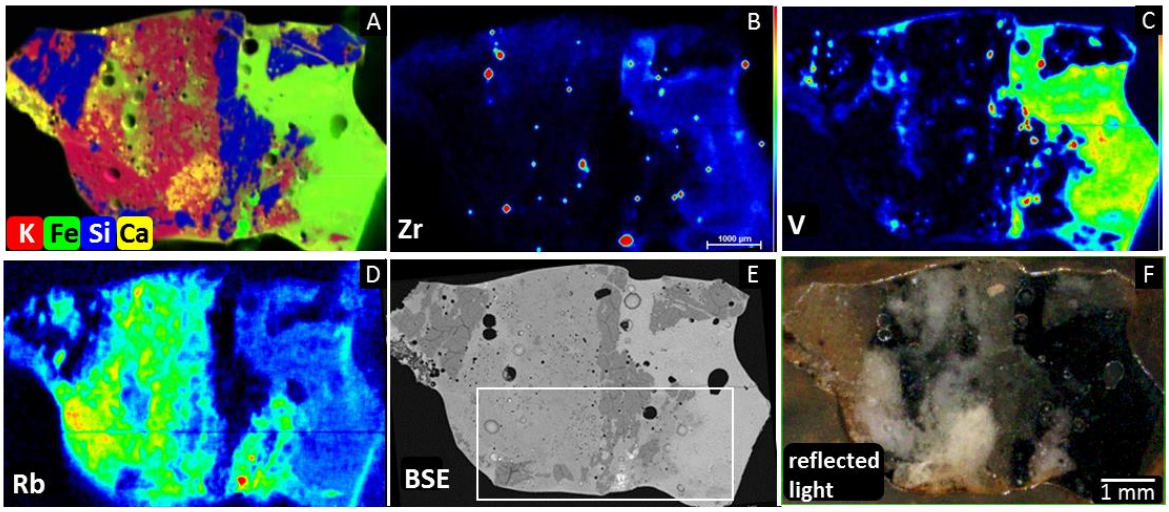


Figure 2. Micro-XRF elemental and electron imagery of vitreous wall fragment. A) Composite x-ray overlay (K K_α-red, Fe K_α-green, Si K_α-blue, Ca K_α-yellow). B-D) single element images in rainbow contrast scale Zr K_α, Rb K_α, and V K_α. E) Backscattered electron image (BSE) mosaic, highlighted area enlarge below. F) Reflected light color image.

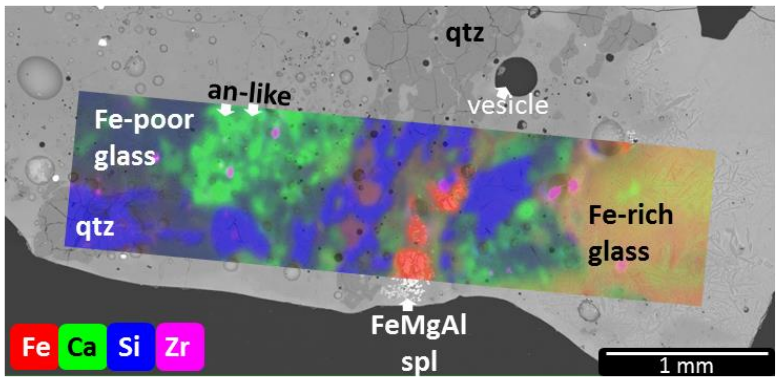


Figure 3. Micro-XRF composite x-ray image superimposed on a BSE image showing the boundary between Fe-rich and -poor glasses [Fe K_α-red, Ca K_α-green, Si K_α-blue, and Zr K_α-majenta; quartz (qtz), anorthite (an), and spinel (spl)].

Table 1. Major and trace element analyses of Fe-rich and –poor glasses. Major elements determined by electron beam x-ray analysis (wt %), and trace elements determined by micro-XRF (ppm).

	Fe-rich glass	Fe-poor glass
SiO_2	54.20	66.65
Al_2O_3	14.69	20.12
CaO	6.42	0.83
MnO	0.31	
FeO*	11.33	
TiO_2	2.26	
MgO	3.73	
Na_2O	3.84	5.18
K_2O	2.15	7.18
P_2O_5	0.71	
Total	99.64	99.95
V	279	
Cr	27	
Mn		122
Fe		1962
Co	781	
Ni		20
Zn	362	
Ga	135	
Rb	152	269
Sr	972	449
Y	203	
Zr	1974	
Ba	86	1517

Weaver, Jamie; Wight, Scott; Pearce, Carolyn; McCloy, John; Lam, Thomas; Sjoblom, Rolf; Peeler, David; Schweiger, Michael; Kruger, Albert; Whittaker, Scott; Vicenzi, Edward.
"Compositional Imaging and Analysis of Late Iron Age Glass from the Broborg Vitrified Hillfort, Sweden."
Paper presented at Microscopy & Microanalysis 2018, Baltimore, MD, United States. August 5, 2018 - August 9, 2018.

SiC Power MOSFET Gate Oxide Breakdown Reliability – Current Status

Kin P. Cheung
Engineering Physics Division
National Institute of Standard & Technology
Gaithersburg, MD USA
Kin.cheung@nist.gov

***Abstract*—The SiC power MOSFET (metal-oxide semiconductor field-effect transistor) is poised to take off commercially. Gate oxide breakdown reliability is an important obstacle standing in the way. Early predictions of poor intrinsic reliability comparing to silicon MOSFET, while theoretically sound, have now proven way too pessimistic. Experimental data from a good quality SiC MOSCAP (MOS Capacitor) turns out to have better breakdown lifetime than its silicon counterpart, based on data available in the literature. This surprising result is the consequence of improper extraction of intrinsic lifetime in the presence of extrinsic failures. Even though intrinsic lifetime is no longer an issue, SiC MOSFET gate oxide breakdown reliability is not out of the woods yet. This is because, for thick oxide, it is the extrinsic failure that determines lifetime, not intrinsic failure. Unfortunately, up to now there is no properly done study of SiC gate oxide extrinsic breakdown reported in the literature.**

***Index Terms*—SiC, MOSFET, Thick Gate Oxide, TDDB, Extrinsic Failures**

I Introduction

The SiC power MOSFET (metal-oxide semiconductor field-effect transistor) is rapidly reaching wide-spread adoption. Reliability issues such as gate oxide breakdown and bias-temperature-instability are largely under control, or apparently so. On the gate oxide breakdown front, while there are lingering extrinsic (early) failure concerns [1], the focus has been on intrinsic reliability and the question is fully settled – there is no intrinsic issue at all [2]. Given how poor the early results were [3], [4] and the theoretically sound argument of limited breakdown lifetime based on the "low" barrier height for tunneling injection of electrons between SiO_2 and SiC [5] as well as the experimentally measured, lower than expected, barrier heights [6], [7], the 'no intrinsic issue' conclusion is truly remarkable.

II Breakdown Data

From the poor results of early days, improvements were reported over time [8-10] so the excellent result reported in [2] is not unexpected. However, it begs the question: what is wrong with the theoretical concern raised in [5] that low barrier height leads to shorter breakdown lifetime? Adding to the mystery is the "unexpected" further lowering of the effective barrier height at the high temperatures at which the SiC MOSFET is expected to operate [6], [7]. It is well known that gate oxide breakdown lifetime is shortened by the current flow across the oxide layer during stress [11-13] and lower tunneling barrier means more current. The concern raised is therefore clearly valid. The reported [6], [7] further lowering of barrier height at high temperature is also consistent with the literature [14]. Thus, theoretically, the gate oxide breakdown lifetime for SiC MOSFET should be far shorter than for silicon MOSFET. The only way to conclude that there is no intrinsic issue with the breakdown reliability of SiC MOSFET gate oxide is that the lifetime of the gate oxide for silicon MOSFET is so extremely long that a drastic reduction still leaves ample room for reliability. However, this seems not to be true.

Figure 1 compares the time-dependent-dielectric-breakdown (TDDB) results of gate oxide for SiC metal-oxide-semiconductor-capacitor (MOSCAP) of [2] to silicon MOSCAP data extracted from literature [15], [16]. Surprisingly, the lifetime of the gate oxide in SiC MOSCAP is far better than the gate oxide in a silicon MOSCAP (MOS Capacitor).

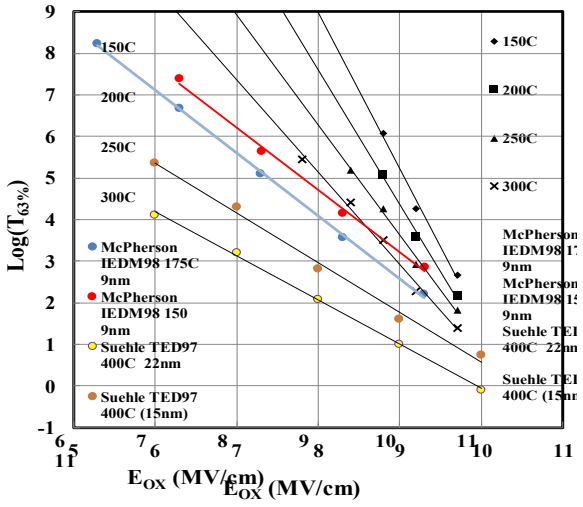


Figure 1 Extracted characteristic failure time (Time to 63 % failure, or τ_{63}) from SiC MOSCAP as a function of stress electric field and temperature. Similar data extracted from literature [15], [16] for silicon is shown for comparison.

How is this possible? The likely answer is that the "intrinsic" lifetimes reported in the literature are not truly intrinsic. Indeed, even the data in [2] may not be truly intrinsic.

Cheung, Kin.
"SiC power MOSFET gate oxide breakdown reliability - Current status."
Paper presented at 2018 IEEE International Reliability Physics Symposium (IRPS), Burlingame, CA, United States. March 11, 2018 - March 15, 2018.

III THE ROLE OF EXTRINSIC FAILURES

No process is perfect. Thus, it is impossible to have no extrinsic failures. However, one can be easily fooled by the absence of early failures in experiments involving a limited number of devices. The literature is full of reports that use 20 or fewer devices in experiments and use the lack of distinct early failure as an indication of the measured distribution being intrinsic and to extract an "intrinsic" characteristic lifetime (τ_{63}). Worse yet, it is not uncommon to see reports that simply ignore the early failure part of the distribution and claim the rest of the distribution as being intrinsic and extract the intrinsic τ_{63} from it. Such practices are highly error-prone. Figure 2 uses two examples to illustrate this point [17].

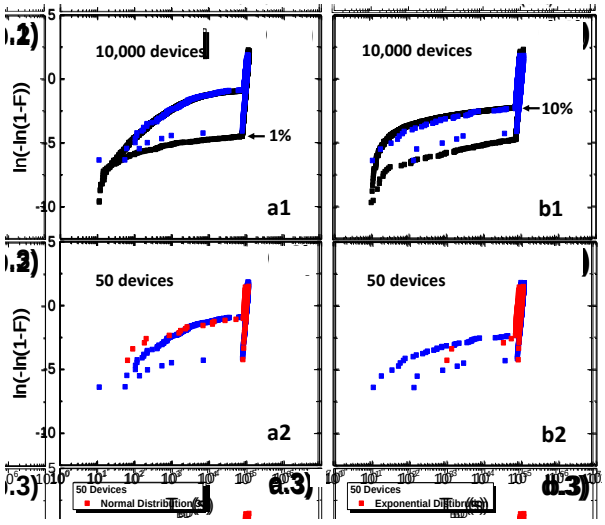


Figure 2 a1 and b1 are two TDDB distributions with different percentage of extrinsic failures. A2 and b2 are the possible corresponding distributions when random sub-sampling of the larger distribution is done.

Figs. 2a1 and 2b1 are two breakdown failure Weibull distributions with clear and significant extrinsic failure tails. Fig. 2a2 is a random sub-sampling of the a1 distribution. Even with 50 devices, the a2 population contains zero early failures. Similarly, fig. 2b2 is a random sub-sampling of the b1 distribution. Even when the parent population extrinsic percentage is very high, the 50 devices population can tempt one to extract an "intrinsic" τ_{63}.

A few groups pointed out that the proper way to treat breakdown distributions with extrinsic failures is to treat it as a joint probability distribution [18-20] of the intrinsic and extrinsic populations and analyze it accordingly. Otherwise, the extract characteristics will have significant errors. Fig. 3 takes the process in reverse to illustrate this point. The blue distribution in fig. 3a is the joint distribution of the two red (straight lines) distributions. One distribution has a large τ_{63} and a β (Weibull shape factor) of 15, representing an intrinsic breakdown distribution, and the other has a small τ_{63} and a small β of 0.5, representing an extrinsic breakdown distribution. The region enclosed by the dotted square is expanded in fig. 3b. In this example, using the steep region to extract τ_{63} will underestimate the intrinsic τ_{63} by roughly 9000 s, or nearly 20 %. Obviously, the size of the error depends on many factors such as the shape factors and the size of the extrinsic fraction. Intuitively, it is clear that the smaller the extrinsic population, the smaller the error. When the number of devices tested is 20 or less, the error is difficult to assess. Since physics dictates that the breakdown lifetime for gate oxides in SiC MOS should be shorter than those in silicon, one can only surmise that the reason for the surprising result of fig. 1 is due to the difference in extrinsic failure population. In [2], 750 devices were tested and no extrinsic failures were observed. The extract lifetime is therefore much closer to the real intrinsic lifetime than those extracted in [15] and [16] with very limited devices.

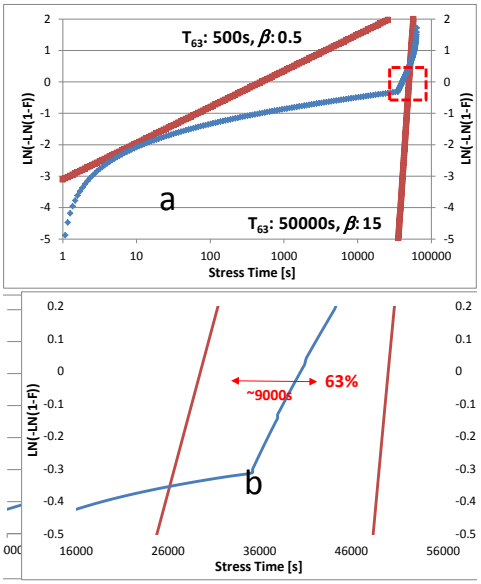


Figure 3 a: shows the joint distribution (blue) resulting from two different distributions, one intrinsic and one extrinsic; b: expanded view of the box (broken line) region showing the significant difference of extracted τ_{63}.

TDDB Data reported in [2] involved 750 MOSCAP devices of the size 0.001 cm^2. This is the first time such a large number of relatively large-size devices were tested without early failures for SiC. Indeed, to the best of my knowledge, no such data has ever been reported for thick gate oxide even for silicon – hence the surprise shown in fig. 1. While this encouraging result clearly dispelled the intrinsic reliability worry, it does not mean that SiC

Cheung, Kin.
"SiC power MOSFET gate oxide breakdown reliability - Current status."
Paper presented at 2018 IEEE International Reliability Physics Symposium (IRPS), Burlingame, CA, United States. March 11, 2018 - March 15, 2018.

MOSFETs have no more gate oxide breakdown reliability concerns.

It is well-established that for thick gate oxide, breakdown reliability is determined by extrinsic failures, not the intrinsic lifetime [18]-[20]. While SiC MOSFETs are now available commercially and impressive breakdown reliability specifications are listed in product data sheets, there are no published results in the literature that suggest the reliability projection is properly done.

A typical TDDB distribution for SiC MOSFET gate oxide has an extensive extrinsic tail similar to figure 4. Depending on the device size and test condition, the extrinsic failure fraction can be larger or smaller. As fig. 2 indicated, the extension of the extrinsic tail will not be clear until enough number of devices are tested. To project product reliability properly, a number of test conditions (electric field and temperature combinations) must be performed and each must have enough number of test devices to ensure the extrinsic tail can be analyzed properly. This means a very large number of devices must be tested. As TDDB is a long-term reliability test that too much acceleration can also lead to errors, this test burden is not easy to meet.

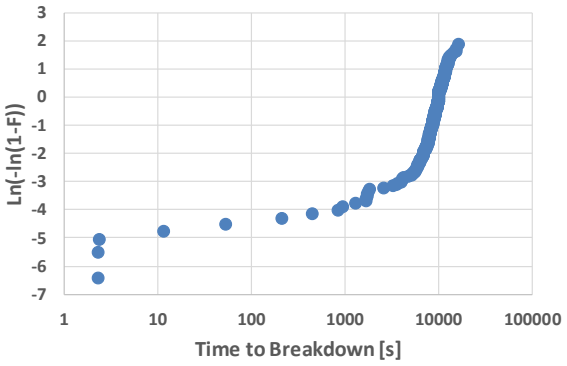


Figure 4 A typical TDDB distribution of SiC MOSCAP showing an extensive extrinsic tail.

It is important to note that the extrinsic failure distribution has a shape factor β often less than one. This means the breakdown lifetime is very area sensitive. As power modules typically involve a large SiC active area, proper area scaling must be performed when projecting product lifetime [21]:

$$\tau_{63}^{P} = \tau_{63}^{T}\left(\frac{A_T}{A_P}\right)^{1/\beta} \qquad (1)$$

where superscript P and T of τ_{63} signifies product and test structure respectively; A_T and A_P are the areas of the test device's active area and product active area. With a large β like those expected from intrinsic failure of thick gate oxide, the characteristic lifetime for product is quite similar to the test structure nearly independent of the area difference. For extrinsic failures where $\beta < 1$, the characteristic lifetime is extremely sensitive to the area ratio. For example, if the test structure is 100 times smaller than the product in terms of active area and the shape factor is 0.5, the characteristic failure time of the product will be 10,000 times shorter than the test structure.

Furthermore, typical reliability requirement is not 63 % failure in 10 years. Instead, it may be 1 % failure or even less in 10 years or longer, depending on the application. The small shape factor's role here is also very large [21]:

$$\tau_{F1} = \tau_{F2}\left(\frac{\ln(1-F_1)}{\ln(1-F_2)}\right)^{1/\beta} \qquad (2)$$

where τ_{F1} and τ_{F2} are time to failure fraction F_1 and F_2 respectively. Translating the 63 % failure time to 1 % failure time with $\beta = 0.5$ will mean dividing the τ_{63} by 10,000, or the time to reach 1 % failure is 10,000 times shorter than to reach 63 % failure.

While, for the same oxide thickness, the intrinsic failure distribution shape factor is independent of the stress electric field, it has been shown that the shape factor for extrinsic failure distributions is dependent on the stress field [19], [20]. This complicates the procedure for lifetime projection and requires the stress test be carried out close to the use field. The good news is that the shape factor seems to increase with decreasing stress field [19], [20] so that when the stress test is not carried out at use field, the lifetime projection is still dependable, albeit a little pessimistic. Whether this is applicable to all thick gate oxide cases are not clear because of the absence of physical understanding of how extrinsic failures occur.

Data in [2], as mentioned before, are rare. Many manufacturers are struggling with significant extrinsic failure population, and relentless effort in improving the cleanliness of wafer processing produced very limited success [22]. There has been no report on using large data set with extrinsic failures to properly project lifetime for SiC MOSFET. It is safe to say that the breakdown reliability issue has not yet been resolved.

IV THE "LUCKY DEFECT MODEL"

Historically, extrinsic breakdown failures in thick gate oxide breakdown are modeled as local-thinning due to contamination of the wafer surface before gate oxide growth [23]-[26]. As a result, the only "trick" to fight extrinsic failure is to improve wafer cleanliness. On the other hand, it was reported in the silicon MOSFET literature that the local-thinning model failed to explain the result of tests with a large sample size [18]-[20]. Furthermore, a local-thinning model based screening method [26], [27] also failed to improve the breakdown distribution [18], [20]. These issues unfortunately fully manifested

Cheung, Kin.
"SiC power MOSFET gate oxide breakdown reliability - Current status."
Paper presented at 2018 IEEE International Reliability Physics Symposium (IRPS), Burlingame, CA, United States. March 11, 2018 - March 15, 2018.

themselves in SiC MOSFET, and the silicon MOSFET literature provided no insight on how to solve them. To address this acute issue, the "Lucky Defect Model" was introduced recently [1].

The "Lucky Defect Model" posits that point defects grown into the gate oxide locally enhance trap-assisted-tunneling (TAT), leading to higher tunneling current at local spots. As higher tunneling current means shorter breakdown time, extrinsic, or early failures resulted. The simulation showed that some commonly assumed point defect distributions can produce the observed extrinsic failure tails. This model points to a new avenue to approach the extrinsic failure problem, which is to improve the oxide growth process, as well as keep the wafer as clean as possible. Data in [2] is the result of extensive process optimization [28].

Figure 5 shows the basic idea of the "Lucky Defect Model." Fig. 5a depicts the case of defect-free oxide with uniform Fowler-Nordheim tunneling at a high applied electric field, in both band diagram and in real space. Fig. 5b depicts, in the real space picture, the case of two point-defects in the oxide layer not far from the injection interface. Each point defect can cause TAT locally as shown in the band diagram. Very similar to the local-thinning model, the shorter breakdown time is due to the locally increased tunneling current. However, unlike the local-thinning model, the source of the current enhancement is not contamination and is not always at the interface.

As TAT is a two-step tunneling process and there exists an optimum distance of the point defect's location from the interface for each applied electric field, the "Lucky Defect Model" predicts the extrinsic tail to be defect profile dependent as well as applied field dependent. This may explain that the shape factor of the extrinsic failure distribution is applied field dependent [18], [20].

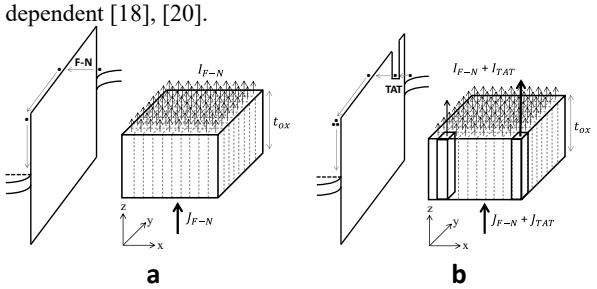


Figure 5a illustrates in real space and in band diagram the case of defect-free oxide Fowler-Nordheim tunneling of current through the oxide layer under high applied electric field. Fig. 5b illustrates the localized increase in tunneling current due to two point-defects in the oxide layer near the injection interface, through trap-assisted-tunneling.

The "Lucky Defect Model" is effective only for thick gate oxides. The reason can be gleaned from (1) which relates the area scaling of lifetime as a function of the Weibull Shape factor β. If we assume the cross-section of the point defect as 1 nm^2 (much smaller than this will make the wavefunction too energetic to be inside the potential well shown in fig. 5b), the impact of this shorter life area on the larger MOSCAP can be calculated. The result is shown in figure 6 for a 0.001 cm^2 MOSCAP. What it shows is the usual smaller area capacitor takes longer to break down and the effect of the Weibull shape factor. If the local enhanced tunneling shortens the local area's lifetime by, say, 10,000 times, the effect will not be detectable if the area scaling makes the smaller area have a lifetime 10,000 times longer. From fig. 6 this happens at a Weibull shape factor just below 3, which corresponds to a gate oxide thinner than 4 nm [29]. Above that thickness, the impact of the shorter life will be measurable. The thicker the oxide, the larger the β and the bigger the impact.

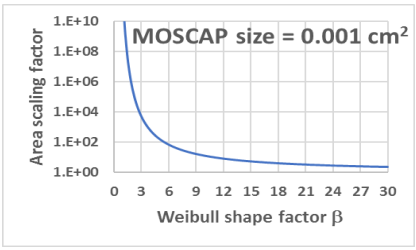


Figure 6 Lifetime area scaling according to (1) for 1 nm^2 defect in a 0.001 cm^2 MOSCAP as a function of the Weibull shape factor which is a function of oxide thickness.

Those who are familiar with thick gate oxide breakdown literature may object to the fact that the "Lucky Defect Model" depends on the idea of the accumulation of defects during stress until a critical value is reached for breakdown to occur. For thick oxides such as those used in SiC MOSFETs, the often-cited model is the feedback runaway model [30]. However, as the thickness of the oxide decreases, the breakdown mechanism transitions to the defect accumulation model. The problem is that the boundary of transition has never been clear. Fortunately, the shape of the leakage current versus time curve during stress provides a clear delineation of the two models. For the breakdown runaway model, the current continues to increase under a constant voltage stress due to hole self-trapping until breakdown occurs [30]. For the defect accumulation model, the current increases at first due to hole self-trapping, then turn around to head lower over time when hole self-trapping saturates [31], [32] and electron-trapping continues to increase. This is clearly demonstrated for the gate oxide in SiC MOSCAP [1]. The "Lucky Defect Model" is already proven useful by pointing to an alternate way to fight extrinsic failures in thick gate oxides [2]. A recent report of excellent gate oxide reliability with very low extrinsic failure also credits improvement in the oxide growth process [33]. Note that while

Cheung, Kin.
"SiC power MOSFET gate oxide breakdown reliability - Current status."
Paper presented at 2018 IEEE International Reliability Physics Symposium (IRPS), Burlingame, CA, United States. March 11, 2018 - March 15, 2018.

the reported result is very impressive with a large area and a large number of devices, it is based on a ramp voltage test only, not TDDB. One can conclude that there is evidence of low extrinsic failures, but how low is not an answer one can ascertain from ramp voltage tests.

V SUMMARY

SiC power MOSFET gate oxide reliability has come a long way. The initial concern about poor intrinsic reliability due to low tunneling barrier is proven overly pessimistic. While the physics is correct, the shorter intrinsic lifetime is still more than adequate. The real issue is extrinsic failure, or early failures due to contamination as well as defects in the oxide layer. For thick oxide, reliability is entirely determined by the extrinsic failure distribution. This requires the testing of a large number of devices with large active areas and proper analysis of the test data. This, unfortunately has not yet been done. As a result, gate oxide breakdown reliability is still a serious problem for SiC power MOSFETs.

REFERENCES

[1] Z. Chbili, A. Matsuda, J. Chbili, J.T. Ryan, J.P. Campbell, M. Lahbabi, D.E. Ioannou, K.P. Cheung, "Modeling Early Breakdown Failures of Gate Oxide in SiC Power MOSFETs," IEEE Trans. on Electron Dev., **63**, 3605-3613 (2016).

[2] Z. Chbili, K.P. Cheung, J.P. Campbell, J. Chbili, M. Lahbabi, D.E. Ioannou, and K. Matocha, "Time Dependent Dielectric Breakdown in high quality SiC MOS Capacitors", Materials Science Forum, **858**, pp. 615-618 (2016).

[3] L. A. Lipkin and J. W. Palmour, "Insulator investigation on SiC for improved reliability," *IEEE Trans. Electron Dev.*, **46**(3), 525–532 (1999).

[4] M. M. Maranowski and J. A. Cooper, Jr., "Time-dependent-Dielectric-breakdown measurements of thermal oxides on n-type 6H-SiC," *IEEE Trans. Electron Devices*, **46**(3), 520–524(1999).

[5] R. Singh and A. R. Hefner, "Reliability of SiC MOS devices," *Solid State Electron.*, **48**(10/11), 1717–1720(2004).

[6] A. K. Agarwal, S. Seshadri, and L. B. Rowland, "Temperature Dependence of Fowler–Nordheim current in 6H- and 4H-SiC MOS capacitors," *IEEE Electron Dev. Lett.*, **18**(12), 592–594(1997).

[7] R. Waters and B. Van Zeghbroeck, "Temperature-dependent tunneling through thermally grown SiO2 on n-type 4H- and 6H-SiC," *Appl. Phys. Lett.*, **76**(8), 1039–1041(2000).

[8] K. Matocha and R. Beaupre, "Time-dependent dielectric breakdown of thermal oxides on 4H-SiC," Mater. Sci. Forum, **556/557**, 675–678 (2007).

[9] L. Yu, K. P. Cheung, J. Campbell, J. S. Suehle, and S. Kuang, "Oxide reliability of SiC MOS devices," in Proc. IEEE Int. IRW, 2008, pp. 141–144.

[10] K. Fujihira, N. Miura, K. Shiozawa, M. Imaizumi, K. Ohtsuka, and T. Takami, "Successful enhancement of lifetime for SiO2 on 4H-SiC By N2O anneal," IEEE Electron Dev. Lett., **25**(11), 734–736(2004).

[11] Chen, I.C., S. Holland, and C. Hu, Oxide breakdown dependence on thickness and hole current-enhanced reliability of ultra-thin oxide. in IEDM. 1986. p. 660.

[12] M. A. Alam, J. Bude, and A. Ghetti, "Field Acceleration for Oxide Breakdown - Can an Accurate Anode Hole Injection Model Resolve the E vs. 1/E Controversy?" IEEE Int. Reliab. Phys. Symp., San Jose, CA, 2000, pp21-26.

[13] K. P. Cheung, "Unifying the thermal–chemical and anode-hole-injection gate-oxide breakdown models," Microelectronics Reliab., **41**(2), 193-199(2001).

[14] Pananakakis, G., G. Ghibaudo, R. Kies, and C. Papadas, "Temperature dependence of the Fowler–Nordheim current in metal-oxide-degenerate semiconductor structures." J. Appl. Phys. **78**(4): 2635-2641(1995).

[15] J. McPherson, V. Reddy, K. Banerjee, and H. Le , "Comparison of E and 1/E TDDB Models fior Si02 under long-termhowfield test conditions," IEDM 1998, pp171-174.

[16] J. S. Suehle and P. Chaparala, "Low electric field breakdown of thin SiO2 films under static and dynamic stress," IEEE Trans. Electron Dev., **44**(5), 801-808(1997).

[17] J. Chbili, Z. Chbili, A. Matsuda4, K. P. Cheung, J. T. Ryan, J. P. Campbell, M. Lahbabi, "Influence of lucky defects distributions on Early TDDB Failures in SiC Power MOSFETs," Accepted for presentation, Int. Integrated Reliability Workshop, Oct. 2017.

[18] Sichart, K. V. and R. P. Vollertsen, "Bimodal lifetime distributions of dielectrics for integrated circuits." Qual. Reliab. Eng. Int., 7(4): 299-305(1991).

[19] R. Degraeve, J. L. Ogier, R. Bellens, P. J. Roussel, G. Groeseneken, and H. E. Maes, "A new model for the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown," IEEE Trans. Electron Dev., **45**(2), 472-481(1998).

[20] Oussalah, S. and B. Djezzar, "Field Acceleration Model for TDDB: Still a Valid Tool to Study the Reliability of Thick SiO2-Based Dielectric Layers?" IEEE TRANS. ELECTRON DEV. **54**(7): 1713-1717(2007).

[21] Wu, E. Y. and R.-P. Vollertsen, "On the Weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination. Part I: theory, methodology, experimental techniques." IEEE Trans. Electron Dev., **49**(12): 2131-2140(2002).

[22] Private communication from more than one manufacturers.

[23] Honda, K., A. Ohsawa, and Nobuo Toyokura, "Breakdown in silicon oxides—correlation with Cu precipitates." Appl. Phys. Lett. **45**(3): 270-271(1984).

[24] Honda, K., A. Ohsawa, and Nobuo Toyokura, "Breakdown in silicon oxides (II)—correlation with Fe precipitates." Appl. Phys. Lett. **46**(6): 582-584(1985).

[25] Wendt, H., H. Cerva, V. Lehmann and W. Pamler, "Impact of copper contamination on the quality of silicon oxides." J. Appl. Phys. **65**(6): 2402-2405(1989).

[26] Lee, J. C., I.-C. Chen, and Chenming Hu, "Modeling and characterization of gate oxide reliability." IEEE Trans. Electron Dev. **35**(12): 2268-2278(1988).

[27] Moazzami, R. and C. Hu, "Projecting gate oxide reliability and optimizing reliability screens." IEEE Trans. Electron Dev. **37**(7): 1643-1650(1990).

[28] Private communication from one manufacturer.

[29] Degraeve, R., G. Groeseneken, R. Bellens, M. Depas, H. E. Maes, "A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides.) IEEE Int. Electron Devices Meeting, 1995, pp863-866.

[30] Ih-Chin, C., S. E. Holland, and Chenming Hu, "Electrical breakdown in thin gate and tunneling oxides." IEEE Trans. Electron Dev. **32**(2), 413-422(1985).

[31] Nissan-Cohen, Y., J. Shappir, D. Frohman-Bentchkowsky, "High-field and current-induced positive charge in thermal SiO2 layers." J. Appl. Phys., **57**(8), 2830-2839(1985).

[32] Nissan-Cohen, Y., J. Shappir, D. Frohman-Bentchkowsky, "Trap generation and occupation dynamics in SiO2 under charge injection stress." J. Appl. Phys., **60**(6): 2024-2035(1986).

[33] Van Brunt, E., D. J. Lichtenwalner, et al. "Reliability Assessment of a Large Population of 3.3 kV, 45 A 4H-SiC MOSFETs." 29th Int. Symp. Power Semiconductor Devices & ICs, Sapporo, Japan. (2017). pp251-254.

Cheung, Kin.
"SiC power MOSFET gate oxide breakdown reliability - Current status."
Paper presented at 2018 IEEE International Reliability Physics Symposium (IRPS), Burlingame, CA, United States. March 11, 2018 - March 15, 2018.

Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis

Jian Yao, Jeffrey Sherman, Tara Fortier, Thomas Parker, Judah Levine, Joshua Savory, Stefania Romisch, William McGrew, Xiaogang Zhang, Daniele Nicolodi, Robert Fasano, Stephan Schaeffer, Kyle Beloy, and Andrew Ludlow

Time and Frequency Division, National Institute of Standards and Technology (NIST), Boulder, Colorado, USA

Abstract

This paper describes the recent NIST work on incorporating an optical clock into a time scale. We simulate a time scale composed of continuously-operating commercial hydrogen masers and an optical frequency standard that does not operate continuously as a clock. The simulations indicate that to achieve the same performance of a continuously-operating Cs-fountain time scale, it is necessary to run an optical frequency standard 12 min per half a day, or 1 hour per day, or 4 hours per 2.33 day, or 12 hours per week. Following the simulations, a Yb optical clock at NIST was frequently operated during the periods of 2017 March – April and 2017 late October – late December. During this operation, comb-mediated measurements between the Yb clock and a hydrogen maser had durations ranging from a few minutes to a few hours, depending on the experimental arrangements. This paper analyzes these real data preliminarily, and discusses the results. More data are needed to make a more complete assessment.

Key Words

Time scale, optical clock, Cs fountain, hydrogen maser, Kalman filter.

I. Introduction

The Al+ [1] and the Sr/Yb [2-3] frequency standards show a stability of better than 1×10^{-16} at an averaging time of several minutes, and even more stable devices are being developed. However, it is challenging to run an optical frequency standard continuously as a clock, making it difficult to incorporate such a device into a conventional time scale, which is based on time differences. In this paper, we present the results of a simulation of a time scale that can accept a device that provides only occasional frequency data in addition to the more traditional time-differences. We also show the preliminary result of incorporating this type of data into the time scale that is used to realize UTC(NIST).

Section II presents the details of the simulation. Also, we describe a time scale using the Kalman filter that can realize the combination of devices we have simulated. Section III presents the simulation results of a time scale that incorporates an optical clock. In particular, we show how long and how often an optical frequency standard must be run in order to make the time scale comparable to the Cs-fountain time scale. The simulated scale is based on free-running continuously-operating hydrogen masers whose characteristics are derived from the real masers in the NIST clock ensemble. Section IV shows the preliminary result of incorporating a real optical clock, i.e., Yb clock, into the NIST time scale. The preliminary real-data result matches the simulation. Section V concludes this paper.

II. Simulation Procedures

There are three steps in the simulation. First, we simulate a free-running time scale, a Cs fountain, and an optical frequency standard. The time scale algorithm is based on a measurement of the time differences every 720 s, which is the measurement interval that is used in the NIST ensemble. A conventional Kalman filter algorithm is used to estimate the frequency and frequency drift of the free-running time scale by using data from either the optical device or the cesium fountain. The frequency stability of the steered time scale is computed.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

II(a). Clock Simulation

We model the characteristics of a "typical" hydrogen maser by using the data from our clock ensemble. The blue curve in Fig. 1 shows the Modified Allan Deviation of the measured time differences between two such masers. We assume that both masers have the same characteristics, so that the red curve in the figure, which is scaled by $1/\sqrt{2}$ relative to the blue curve, is our estimate of the stability of a single maser. The dotted lines in the figure show our model of the observed stability.

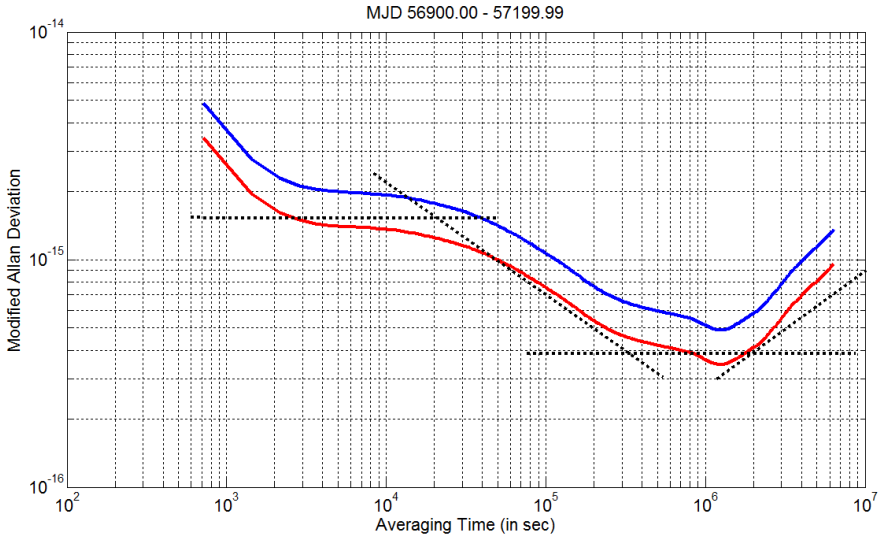


Figure 1. The blue curve shows the frequency stability of the time difference between two NIST H-masers. The red curve shows the estimated frequency stability of a single H-maser, assuming that the two masers have the same statistical characteristics. We model the stability as shown by the black dotted lines.

We generate a time series containing the sum of white frequency noise, flicker frequency noise and random-walk frequency noise with noise parameters chosen so that the time scale, which consists of N identical masers, is more stable than a single H-maser by a factor of $1/\sqrt{N}$. (This improvement is rigorously true for white noise processes where there is no correlation between the noise contributions of each of the masers. Our experience is that it is also true for any noise process in an ensemble with no correlations.) Second, we apply the phenomenologically-determined moving average algorithm in Eq. (1). This algorithm is chosen so as to make the frequency-stability curve flat for averaging times less than or equal to 20 000 s so that the simulation is close to the statistics of our masers. (The white phase noise for averaging times less than 2000 s in Fig. 1. is included in the model for the time difference measurements.) In the following equation, $x(n)$ denotes the generated data value at epoch n and $y(n)$ denotes the output of the averaging process at the same epoch. The Modified Allan Deviation of the simulated maser is shown by Figure 2. An ensemble of N masers would have a Modified Allan Deviation that was smaller by a factor of $1/\sqrt{N}$.

$$y(n) = \frac{1}{3}x(n) + \frac{2}{3}\cdot\frac{1}{3}x(n-1) + \left(\frac{2}{3}\right)^2\cdot\frac{1}{3}x(n-2) + \left(\frac{2}{3}\right)^3\cdot\frac{1}{3}x(n-3) + \cdots \quad (1)$$

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

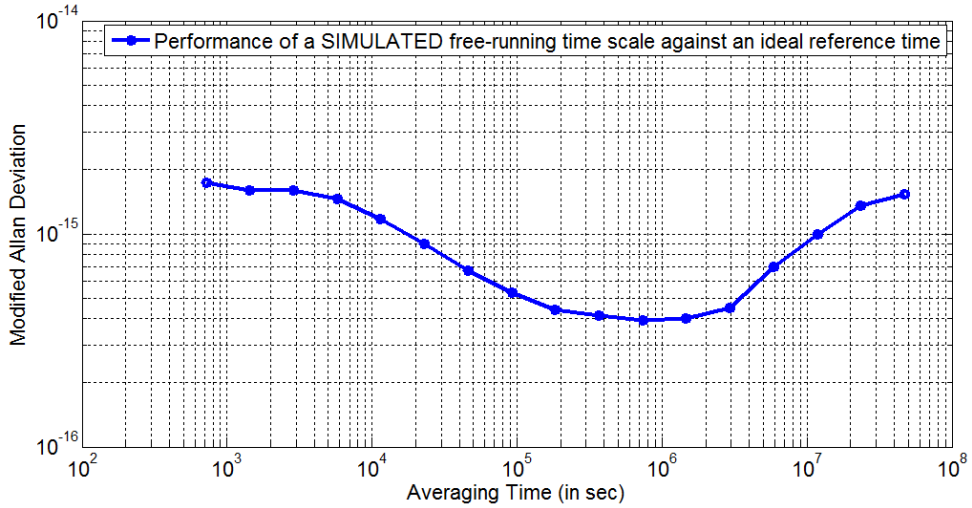


Figure 2. Frequency stability of a simulated free-running time scale with one maser against an ideal reference time.

The continuously-operating cesium fountain is simulated by white frequency noise with a fractional frequency stability of 1.4×10^{-15} at 1 day, which is typical Cs fountain performance at low atom density [4-6]. We simulate an optical clock as having white frequency noise with fractional frequency stability of 3.4×10^{-16} at 1 s, which is typical for the Sr/Yb optical-lattice clock [2-3]. The simulated optical-clock data are truncated to match various operational schedules as discussed below.

II(b). Steering Algorithm

The Kalman filter is used to estimate the frequency and frequency drift of the free-running time scale with respect to the Cs fountain or the optical clock, and the frequency and frequency drift of the time scale are steered based on this estimate.

There are two basic equations in the Kalman filter [7]. Eq. (2) is the system model, which predicts the state of the system at epoch k+1 based on its state at epoch k. Here, $X(k)$ is the estimate state vector of the system at epoch k, and $X(k+1|k)$ is the predicted state vector of the system at epoch k+1. Φ is the transition matrix, which links $X(k)$ and $X(k+1|k)$; it is determined by the physical properties of the system. u is the process noise, which is characterized by the Q matrix. Eq. (3) is the measurement model. The H matrix gives the relation between the state vector X and the measurement vector Z. v is the measurement noise, which is characterized by the R matrix. The R matrix is determined by the details of the measurement process, and its detailed specification is outside of the scope of this simulation. We assume that the white frequency noise of the maser will make a significant contribution to the R matrix, and that other contributions will be much smaller.

$$X(k+1|k) = \Phi \cdot X(k) + u, \quad u \sim N(0, Q). \tag{2}$$

$$Z(k+1) = H \cdot X(k+1|k) + v, \quad v \sim N(0, R). \tag{3}$$

The estimated state vector at epoch k+1 is

$$X(k+1) = X(k+1|k) + K \cdot (Z(k+1) - H \cdot X(k+1|k)), \tag{4}$$

where K is the Kalman gain matrix, which is given by [7, Chapter 4],

$$K = P(k+1|k) \cdot H^T \cdot [H \cdot P(k+1|k) \cdot H^T + R]^{-1}, \tag{5}$$

where $P(k+1|k) = \Phi \cdot P(k) \cdot \Phi^T + Q$, and $P(k+1) = [I - K \cdot H] \cdot P(k+1|k)$.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

For our clock system, Eq. (2) becomes

$$\begin{pmatrix} f_{diff}(k+1|k) \\ d_{diff}(k+1|k) \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} f_{diff}(k) \\ d_{diff}(k) \end{pmatrix} + \begin{pmatrix} \eta \\ \zeta \end{pmatrix}, \tag{6}$$

where f_{diff} is the fractional frequency difference between the time scale and the external frequency standard, d_{diff} is the fractional frequency-drift difference, and Δt is the time interval between epoch k and epoch k+1. The constant time interval is 720 s. The parameters η and ζ are the noise terms in frequency and frequency drift, respectively [8-9]. Since the frequency noise of the external frequency standard is assumed to be negligible, η mainly comes from the frequency noise of the free-running time scale. ζ is typically too small to be observed for a measurement time interval of 720 s. Thus, we set the variance of ζ to a small value. In other words, the model assumes that the frequency drift is essentially constant over the measurement interval of 720 s.

If the Cs fountain is used as the reference, the measurement vector Z is the scalar frequency difference between the time scale and the Cs fountain measured every 720 s. If the optical clock is used as the reference, we do the pre-processing to get the average frequency difference between the time scale and the optical clock during the operation period of the optical clock. (As discussed in the previous paragraph, this assumes that the frequency drift of the maser is constant during this interval.) Then we use the average frequency difference as the measurement vector Z.

The measurement process provides an estimate of frequency, so that the H matrix of the measurement system is

$$H = (1 \quad 0). \tag{7}$$

The frequency and frequency drift at epoch k+1 are given by Eq. 4

$$\begin{pmatrix} f_{diff}(k+1) \\ d_{diff}(k+1) \end{pmatrix} = \begin{pmatrix} f_{diff}(k+1|k) \\ d_{diff}(k+1|k) \end{pmatrix} + K \cdot \left(Z(k+1) - H \cdot \begin{pmatrix} f_{diff}(k+1|k) \\ d_{diff}(k+1|k) \end{pmatrix} \right), \tag{8}$$

where K is the Kalman gain matrix, which can be calculated using Eq. (5).

Finally, we use the updated estimates to steer the time scale by adjusting its frequency and frequency drift.

III. Simulation Results

III(a). Simulated Performance of a Cs-Fountain Time Scale

In this section, we first show the simulated performance of a time scale steered to a continuously-operating cesium fountain. The results of the simulation are shown in fig. 3. The green dashed line is the estimate of the stability of a typical cesium fountain operating at low atom density. The solid curves show the stability of a steered time scale containing one or four masers. The stability of an ensemble of N masers would be scaled by a factor of $1/\sqrt{N}$ relative to the one-maser values.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

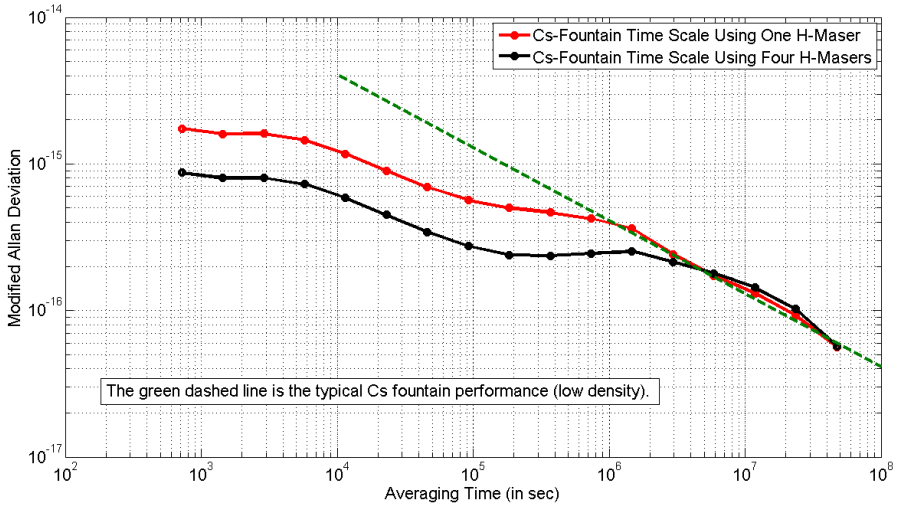


Figure 3. Performance of a time scale steered to a continuously-operating cesium fountain. The green dashed line shows the performance of the cesium fountain. The solid curves show the stability of a steered time scale that contains either one or four masers.

All of the ensembles are more stable than the cesium fountain standard for averaging times less than about 10 days. The figure illustrates the statistical advantage of steering an ensemble of multiple masers rather than a single device. This point is discussed in more detail in section V. This advantage is in addition to the ability of an ensemble algorithm to minimize the impact of the failure of any one of the masers and to provide a flywheel should the link to the cesium fountain fail for any reason.

A real-world Cs-fountain time scale, such as UTC(OP) [6] and UTC(PTB) [5], has a quite similar performance to the red curve in Figure 3. For example, analyzing the UTC(OP) data during Modified Julian Date 56649 – 57514 indicates that the frequency stability of UTC(OP) is about 4.8×10^{-16} at 10 days, and 2.4×10^{-16} at 60 days. The small difference between UTC(OP) and our simulation may come from the time-transfer noise and the fact that the Cs fountain does not run 100 % of the time.

III(b). Simulated Performance of a Time Scale with an Intermittently-Operating Optical Clock

In Figure 4, the orange dotted line is the modeled stability of a Sr/Yb optical clock, with a stability dominated by white frequency noise with an amplitude of 3.4×10^{-16} at 1 s. The green dotted line is the estimate of the stability of a continuously-operating cesium fountain as in the previous figures. The solid curves show the stability of a time scale ensemble of one maser when its frequency and frequency drift are estimated and corrected from the optical clock data.

In the upper left plot of Figure 4, the optical clock runs every half a day for either 12 or 24 minutes. In both of these cases, we can achieve the same performance as that of a continuously-operating Cs fountain time scale. The operation time of the optical clock is only 1.66 %, for the blue curve. We can also see that there is little improvement if we double the running time to 24 min every 12 hours. This is due to the maser's flat noise characteristics at times on the order of 1000 s (see Figure 2). The upper right plot shows the simulated performance of the same time scale with an optical clock running for various time intervals once a day. If we run an optical clock 12 hours a day (grey curve), the frequency stability of the time scale will be improved to three times better than the Cs-fountain time scale for an averaging time of greater than 10 days. Even with a 50 % duty cycle, the stability of the steered time scale is significantly worse than the stability of the optical clock itself. The bottom left plot shows the simulated performance of the same time scale with an optical clock running three times every week. To achieve performance similar to that of a Cs-fountain time scale, we need to run an optical clock for at least four hours, three times a week. The bottom right plot shows the simulated performance of the steered time scale with an optical clock running once a week. The

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

optical clock must operate for 12 hours each week to make the optical-clock time scale comparable to a Cs-fountain time scale. Practically, a free-running time scale can be significantly pulled by an abnormal hydrogen maser, and the error becomes larger over time. Thus, a frequent evaluation of the free-running time scale using the optical clock is necessary to avoid a large timing error. From this perspective, the actual performance of a time scale steered once a week is unlikely to be as good as the simulation indicated in the bottom right plot.

Figure 5 shows the improvement that can be realized by steering an ensemble of N masers rather than just a single device as in the previous figures. The optical clock runs for one hour per day in these simulations. The stability has an improvement of $\sqrt{N}$, for all averaging times. Comparing Figure 5 with Figure 3, for an averaging time of greater than 10 days, using an ensemble of masers, instead of a single maser, does not reduce the instability of a Cs-fountain time scale, while there is an improvement of $\sqrt{N}$ for an optical-clock time scale by using an ensemble of masers. Comparing the black curve in Figure 5 with the orange curve of the upper right plot of Figure 4, the time scale composed of four masers and one optical clock running an hour a day is better than the time scale composed of one maser and one optical clock running two hours a day.

In summary, to achieve the same performance as a Cs-fountain time scale, we could run an optical clock according to one of the following options: 12 min per half a day, 1 hour per day, 4 hours per 2.33 day, or 12 hours per week. The results are not sensitive to the assumed stability of the optical clock, provided only that it is much more stable than the free-running time scale.

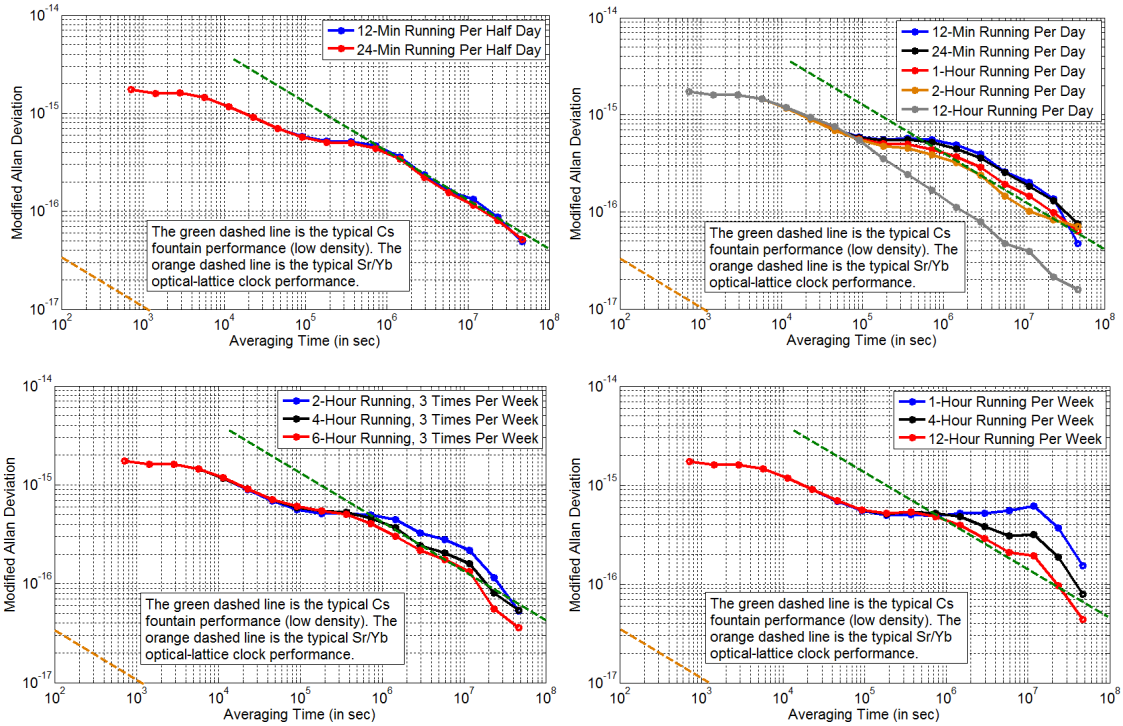


Figure 4. Performance of a time scale steered to an intermittently-operating optical clock. A variety of running schedules are studied in this figure.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

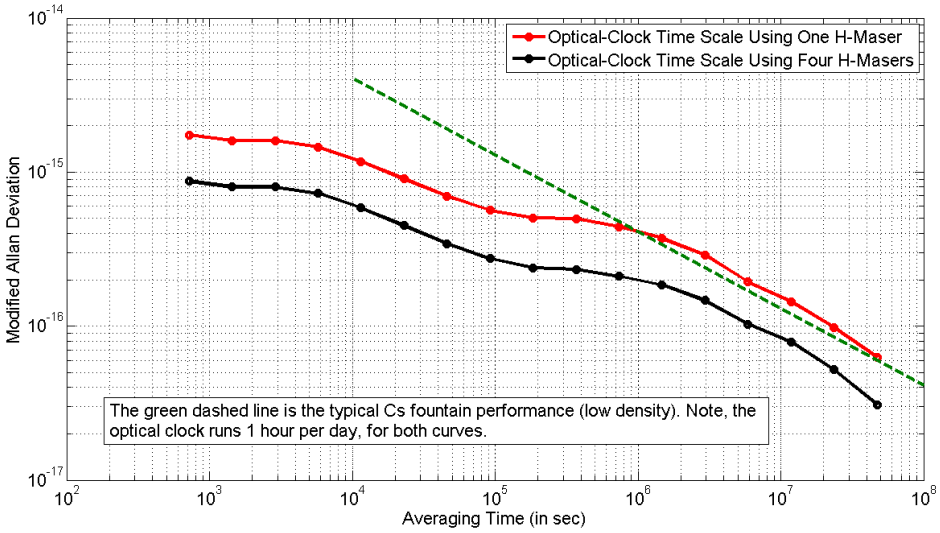


Figure 5. The simulated results of steering an ensemble of masers to an optical frequency reference that runs one hour per day. An ensemble of four masers steered to an optical clock running one hour per day is more stable than a cesium fountain for all averaging times.

IV. Preliminary Real-Data Analysis

IV(a). Initial Campaign in 2017 March and April

Following the simulations, the ytterbium optical clock at NIST [2] has been run regularly in 2017 March and April. Individual run times range from 48 min to 6.9 hours, depending on the experimental arrangement. The Yb frequency standard was down-converted to around 1 GHz via a Ti:sapphire frequency comb. Then this microwave signal was compared with the up-converted hydrogen maser signal, via a mixer (Note, the 5 MHz signal from a hydrogen maser is up-converted to 1 GHz). The frequency difference between the two microwave signals was measured by a commercial frequency counter. The fractional frequency difference between the Yb clock and the hydrogen maser can be derived based on this measurement. In addition, the gravitational correction due to the height from the geoid was applied to the Yb clock so that we know the fractional frequency difference between the "geoid" Yb clock and the hydrogen maser. For simplicity, the Yb clock mentioned below means the "geoid" Yb clock.

Figure 6 shows the fractional frequency difference between Yb and a NIST hydrogen maser ST15. Similar to what we have done in the simulations (see Section II(b)), we average each individual run and compute the mean fractional frequency difference between Yb and ST15. According to the simulation result in Figure 5, an ensemble steered to an optical clock is more stable than a single clock steered to an optical clock, for all averaging times. Thus, to completely exploit the advantages of having an optical clock, we steer the free-running AT1 time scale to the optical clock, instead of steering a single H-maser to the optical clock. Admittedly, in practice, the measurement noise of the link between the H-maser and the time scale could make this architecture noisier. However, this noise is mainly white phase noise and is averaged down to become negligible after tens of minutes. Thus, the architecture of steering the time scale to an optical clock is still more favorable in practice. Figure 7 shows the average fractional frequency difference between AT1 and the Yb clock, for each individual run. The variation between AT1 and Yb is about 5e-15 s/s. The frequency offset shown in Figure 7 mainly comes from the frequency offset of AT1. We can also observe a tiny frequency drift of AT1 from Figure 7. The information of the frequency drifts of AT1 and hydrogen masers based on an optical clock is one of a few benefits of having an optical clock running. This information is helpful for the timekeeping at NIST. The steer of AT1 to Yb is achieved by the Kalman filter as mentioned in Section II(b). An intuitive understanding of this filter is: the longer operation time of the Yb clock, the larger weight the run gets; the longer time interval between

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

the previous run and the current run, the larger weights the current run gets. At the current stage, the steer is done on paper, instead of tuning an AOG (auxiliary output generator) physically.

To evaluate the performance of the steered AT1, we compare the steered AT1 with the UTC via the TWSTFT (two-way satellite time and frequency transfer) link. The result is shown by the left plot of Figure 8. We can see that there is a constant frequency offset of -9.14e-15 s/s between the steered AT1 and the UTC. This offset could come from a variety of sources, e.g. the Yb clock, the frequency comb, the measurement system, or an analysis error. Indeed during these measurements, the Yb clock frequency was being intentionally varied as part of an independent systematic-uncertainty-evaluation, and the frequency counter used to count the comb repetition rate was being operated in a mode where known biases exist. In a more recent campaign (see Section IV(b)), where systems were operated normally and the frequency counting was done in a high accuracy configuration, we observe no such offset. The right plot of Figure 8 shows the residual of the steered AT1 against UTC (red curve), and the residual of the AT1 against UTC (blue curve) as a comparison. The standard deviation of the residual of the steered AT1 is only 0.12 ns, and the corresponding Allan deviation is about 4.4e-16 at 5 days with an upper bound of 7.9e-16 and a lower bound of 3.4e-16. Also, we can see that the frequency drift in AT1 is corrected by steering AT1 to the Yb clock.

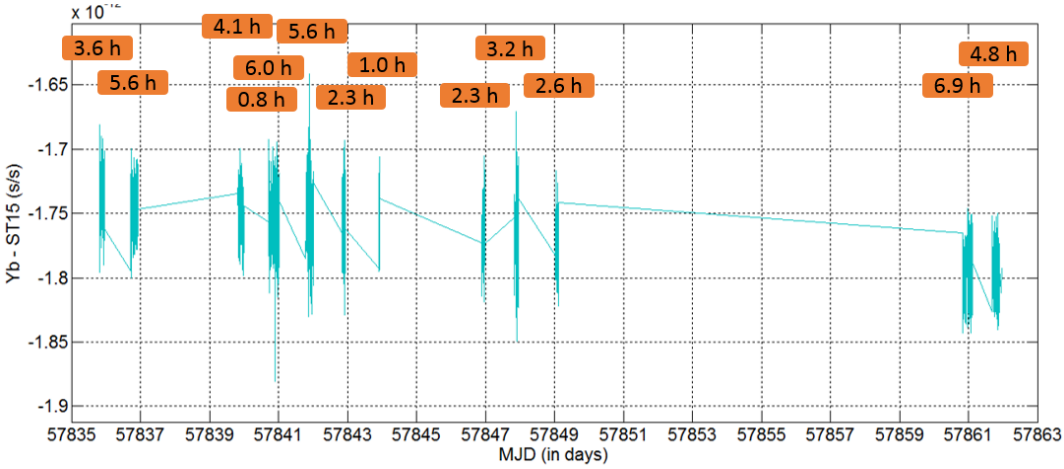


Figure 6. Comparison of the Yb optical clock and a NIST hydrogen maser (i.e., ST15), for the initial campaign.

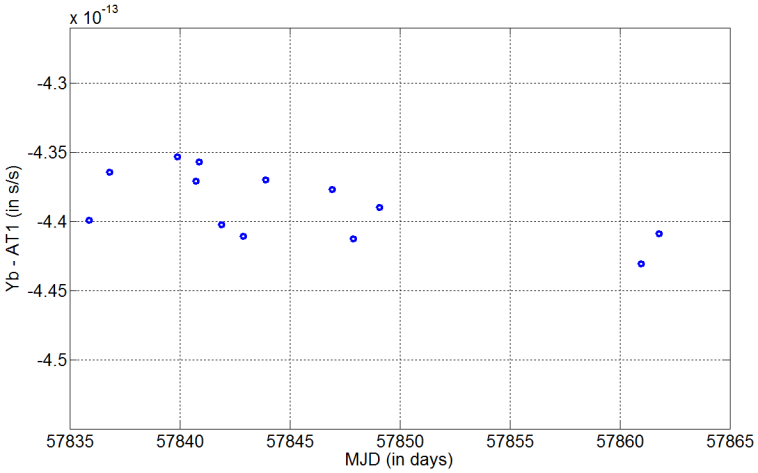


Figure 7. Average of "Yb – AT1," for the initial campaign.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

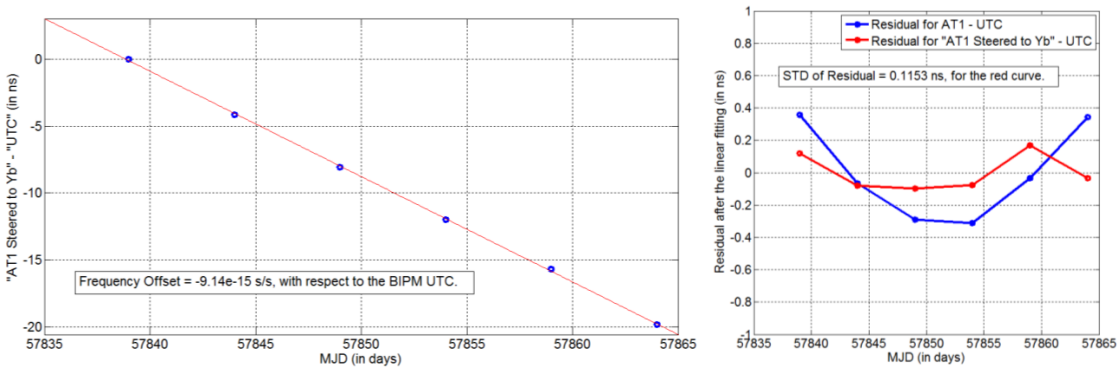


Figure 8. Steered AT1 versus UTC (left) and the residual after the linear fitting (red curve in the right plot), for the initial campaign.

IV(b). Ongoing Campaign since 2017 late October

Encouraged by the above resulting statistical performance, since 2017 late October, we have conducted a focused campaign with the ytterbium clock and optical frequency comb. In particular, the Yb clock is run in normal operation including a formal accounting of systematic clock shifts. Furthermore, frequency offsets in the frequency-comb-based optical-to-microwave synthesis were characterized and confirmed to be small, and the frequency counter was operated in a mode with essentially negligible bias. As mentioned above, during this campaign, no sizeable frequency offset was measured versus UTC. This campaign is ongoing, and the expected duration of more than a few months can provide better statistics than the initial campaign.

Figure 9 shows the data from Oct. 26, 2017 to Dec. 28, 2017. The error bar of each data point mainly depends on the instability of AT1, if we assume that the noise from the Yb clock and the measurement system is negligible. Depending on the time duration of the comparison between Yb and AT1, the error bar could vary from high 10^{-16} to low 10^{-15}. From the data shown, there were two measurements with the optical-clock that lasted for only 7 min, which are labeled by the red ellipses. These two data points have big error bars and should not be considered as outliers. They are given the weights that they deserve, in the Kalman filter. From Figure 9, we observe a fractional frequency change in "Yb – AT1" of about $+1\times10^{-15}$ from the first group of data sets (i.e., MJD 58053 – 58075) to the second group of data sets (i.e., MJD 58083 – 58110). Comparing AT1 with the BIPM UTC cannot resolve any obvious frequency drift in AT1. When comparing the maser-based measurement of Yb clock with the BIPM UTC shown by Figure 10, we see a difference in the two data sets of about $+1\times10^{-15}$.

Over the same time period, another very preliminary independent study has compared the measured Yb frequency with the Cs frequency (i.e., PSFS – Primary and Secondary Frequency Standards) published in the BIPM Circular T. For the first group of data sets, the average frequency difference between Yb and PSFS is -1.2×10^{-15}, with an estimated uncertainty of 5×10^{-16}. This uncertainty includes a dead time uncertainty of 3.7×10^{-16}, a frequency transfer uncertainty of 2.6×10^{-16}, and a PSFS uncertainty of 2.0×10^{-16}. It does not include the u_A and u_B uncertainties for Yb. The inherent uncertainties for Yb should be very small, but there may be an additional u_A uncertainty due to a possible frequency error that may occur in reducing the optical frequency down to the 1 GHz range where it can be measured relative to a hydrogen maser with a counter. This additional uncertainty is currently being evaluated. For the second group of data sets, the average frequency difference between Yb and PSFS is -1×10^{-17}, with an uncertainty of 6×10^{-16} (again not including u_A and u_B). Thus, the two groups of data sets agree with each other at the 1.5 sigma level. Since the campaign is ongoing, more data will be collected to make a more complete assessment.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

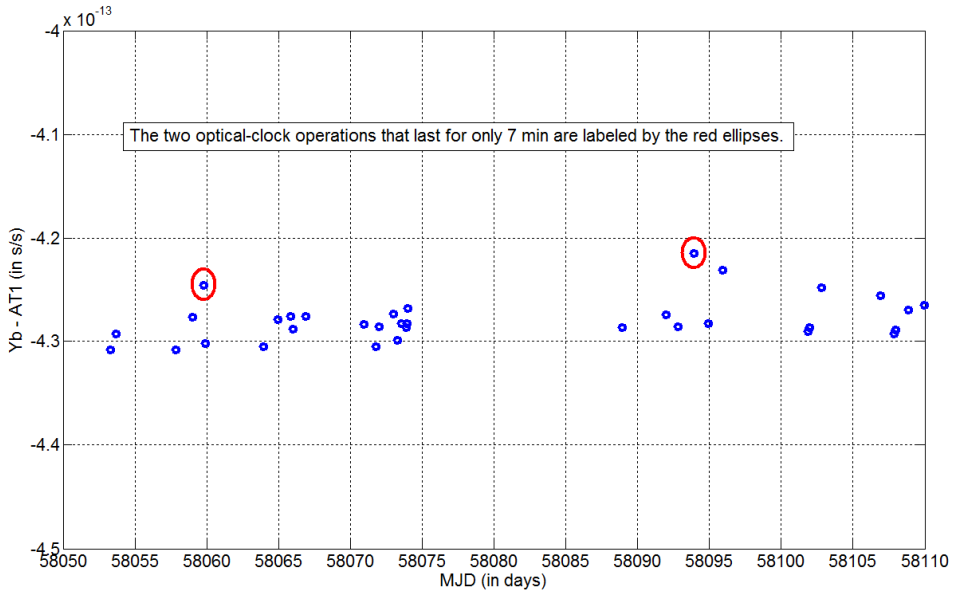


Figure 9. Average of "Yb – AT1," for the ongoing campaign.

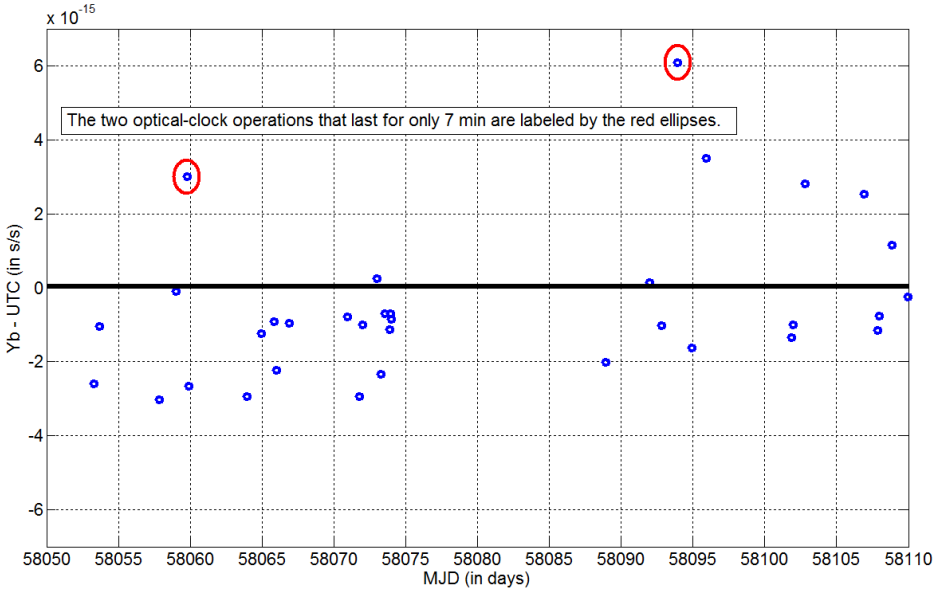


Figure 10. Average of "Yb – UTC," for the ongoing campaign.

V. Future Work

The measurement campaign described in Section IV is ongoing, and with additional data in the coming months, we will know more about the statistics of Yb versus UTC/PSFS. Furthermore, NIST has two other optical clocks which

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

are sometimes operated: the Al+ clock at NIST [1] and a Sr clock at JILA [3]. We plan to explore the scheme of having multiple optical clocks in a future time scale. This can distribute the work to different groups of people and thus ease the work intensity.

Acknowledgments

The clock noise in this paper is generated by the method proposed in [10]. We also thank Chris Oates and Scott Diddams for their helpful inputs. Last, the earlier experimental work done by PTB [11] and NICT [12] inspires our work.

Contributions

Jian Yao did the simulations, designed the Kalman filter, and analyzed the real data. Jeffrey Sherman, Tara Fortier, Andrew Ludlow, William McGrew, Xiaogang Zhang, Daniele Nicolodi, Robert Fasano, Stephan Schaeffer, and Kyle Beloy performed experiments, including the measurement system, the frequency comb, and the Yb optical clock. Other authors assisted in this work.

References

[1] C. W. Chou, D. B. Hume, J. C. J. Koelemeij, D. J. Wineland, and T. Rosenband, "Frequency comparison of two high-accuracy Al^+ optical clocks," Phys. Rev. Lett. 104, 070802, 2010.
[2] N. Hinkley, J. A. Sherman, N. B. Phillips, M. Schioppo, N. D. Lemke, K. Beloy, M. Pizzocaro, C. W. Oates, and A. D. Ludlow, "An atomic clock with 10^{-18} instability," Science, vol. 341, pp. 1215-1218, 2013.
[3] B. J. Bloom, T. L. Nicholson, J. R. Williams, S. L. Campbell, M. Bishof, X. Zhang, W. Zhang, S. L. Bromley, and J. Ye, "An optical lattice clock with accuracy and stability at the 10^{-18} level," Nature, vol. 506, pp 71-75, 2014.
[4] T. P. Heavner, S. R. Jefferts, E. A. Donley, J. H. Shirley, and T. E. Parker, "NIST-F1: recent improvements and accuracy evaluations," Metrologia, vol. 42, pp. 411-422, 2005.
[5] A. Bauch, S. Weyers, D. Piester, E. Staliuniene, and W. Yang, "Generation of UTC(PTB) as a fountain-clock based time scale," Metrologia, vol. 49, pp. 180-188, 2012.
[6] G. D. Rovera, S. Bize, B. Chupin, J. Guena, Ph. Laurent, P. Rosenbusch, P. Uhrich, and M. Abgrall, "UTC(OP) based on LNE-SYRTE atomic fountain primary frequency standards," Metrologia, vol. 53, pp. S81-S88, 2016.
[7] A. Gelb, "Applied optimal estimation," MIT Press, 1974.
[8] L. Galleani, and P. Tavella, "On the use of the Kalman filter in timescales," Metrologia, vol. 40, pp. S326-S334, 2003.
[9] J. Levine, "Invited review article: The statistical modeling of atomic clocks and the design of time scales," Review of Scientific Instruments, vol. 83, 021101, 2012.
[10] N. Ashby, "Probability distributions and confidence intervals for simulated power law noise," IEEE Transactions on UFFC, vol. 62, pp. 116-128, 2015.
[11] C. Grebing, A. Al-Masoudi, S. Dorscher, S. Hafner, V. Gerginov, S. Weyers, B. Lipphardt, F. Riehle, U. Sterr, and C. Lisdat, "Realization of a timescale with an accurate optical lattice clock," Optica, vol. 3, pp. 563-569, 2016.
[12] T. Ido, H. Hachisu, F. Nakagawa, and Y. Hanado, "Time scale steered by an optical clock," Proceedings of the 48th Precise Time and Time Inverval Meeting, pp. 15-17, 2017.

Yao, Jian; Sherman, Jeffrey; Fortier, Tara; Parker, Thomas; Levine, Judah; Savory, Joshua; Romisch, Stefania; McGrew, William; Fasano, Robert; Schaeffer, Stefan; Beloy, Kyle; Ludlow, Andrew.
"Incorporating an Optical Clock into a Time Scale at NIST: Simulations and Preliminary Real-Data Analysis."
Paper presented at ION International Technical Meeting/Precise Time and Time Interval Systems and Applications Meeting (ITM/PTTI 2018) , Reston, VA, United States. January 29, 2018 - February 1, 2018.

Vibrational Interferometry Enables Single-Scan Acquisition of all $\chi^{(3)}$ Multi-Dimensional Coherent Spectral Maps

T. M. Autry[1,*], G. Moody[1], C. McDonald[1,2], J. M. Fraser[1,3], R. P. Mirin[1], K. L. Silverman[1]

[1]*National Institute of Standards and Technology, Boulder CO 80305, USA*

[2]*Department of Physics, University of Colorado, Boulder CO 80309-0390, USA*

[3]*Queen's University, Kingston ON K7L 3N6, Canada*

**e-mail: travis.autry@nist.gov*

Abstract: We demonstrate a new method for multidimensional coherent spectroscopy of nanostructures. We use a heterodyne technique implemented with a confocal microscope to record the amplitude and phase of all degenerate third-order wave-mixing processes.

OCIS codes: (300.6300) Spectroscopy, Fourier transforms; (320.7130) Ultrafast processes in condensed matter, including semiconductors.

1. Introduction

Multidimensional Coherent Spectroscopy (MDCS) is a powerful method for probing the optical properties of physical systems by correlating the absorption, emission, and mixing frequencies of a nonlinear four-wave mixing (FWM) signal. MDCS spectra have been demonstrated to unambiguously measure coherent vs. incoherent coupling as well as homogenous and inhomogeneous linewidths in a variety of systems. Conventional methods for MDCS rely on wave-vector phase matching—a technique that fails for emission from point sources and is not compatible with nano-emitters integrated into waveguide geometries. Alternative methods appropriate for individual nano-objects such as quantum dots, single molecules, or carbon nanotubes rely on phase cycling to detect "action" signals as fluorescence, photoelectrons, or photocurrent. These methods suffer from poor collection efficiency and have limited ability to amplify the signal, since heterodyne detection is not possible. Furthermore, these techniques can only measure FWM signals across a ~10 ps timescale, or they can lack a suitable optical reference to properly process the signals [1-6].

2. Methods

Here, we present a new collinear method, appropriate for long-lived dipole radiation in the far field, that circumvents these limitations through vibrational-interferometry of the inter-pulse delays using a far-detuned continuous-wave reference laser. Our approach uses a pulse-to-pulse phase-cycling scheme with heterodyne detection insuring shot noise-limited signal-noise Unlike conventional methods, we record *all* degenerate wave-mixing processes in a *single* measurement, *i.e.,* zero-quantum and one-quantum FWM signals (both rephasing and non-rephasing pathways), the two-quantum FWM signal, as well as time-resolved linear absorption. Simultaneous measurement of these signals enables complete characterization of the $\chi^{(3)}$ resonant nonlinear optical response of nano-systems with instrument limited ~ 8 orders of temporal dynamic range, ~4 μeV resolution/ 21eV bandwidth, signals as low as 1 photon/pulse, and ~10's of attoseconds precision.

To achieve pulse-to-pulse phase cycling, we drive an acousto-optical modulator (AOM) in each arm of a nested Mach-Zehnder interferometer at radio frequencies that differ in the MHz range (Fig. 1A). The AOMs provide a unique

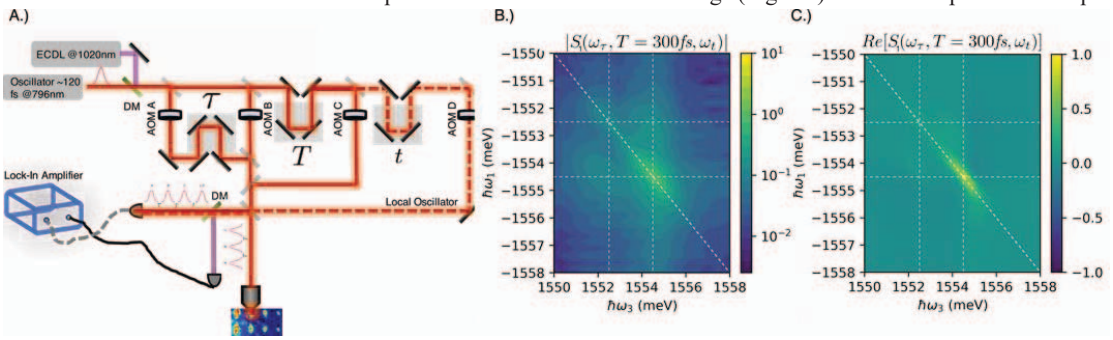


Figure 1 A.) Experimental apparatus showing two Mach-Zehnder interferometers nested within a larger interferometer. Each arm of the interferometer has an AOM driven at a unique radio frequency. At the interferometer entrance/output, both a Ti:Sapph. oscillator (at 796 nm) and a CW laser (at 1020 nm) are combined/split on a dichroic beamsplitter (DM). B.) Measured amplitude of a rephasing pulse sequence showing three orders of magnitude signal-to-noise. C.) Real part of rephasing pulse sequence demonstrating the recovery of the four-wave mixing phase.

Autry, Travis; Moody, Galan; Fraser, James; McDonald, Corey; Mirin, Richard; Silverman, Kevin.
"Vibrational Interferometry Enables Single-Scan Acquisition of all {chi}^u(3)^ Multi-Dimensional Coherent Spectral Maps."
Paper presented at Conference on Lasers and Electro-Optics & Quantum Electronics and Laser Science Conference (CLEO/QELS),
San Jose, CA, United States. May 13, 2018 - May 18, 2018.

carrier-envelope offset for each arm of the interferometer. Experimentally this manifests as a pulse-to-pulse phase accumulation between interferometer arms that evolves at the "beat" frequency of the different AOM drive frequencies. Conveniently, linear signals are recorded at the beat of each pair of interferometer arms, while nonlinear FWM signals are recorded at the higher-order beats involving all four arms of the interferometer. Three pulses (ABC) are recombined and sent through single-mode fiber before focusing through the microscope. A signal is collected in the reflected direction and heterodyne-mixed with a local oscillator through a single-mode photodetector

Unlike conventional methods of achieving phase stability, our interferometer has no common optical path, no requirements for actively stabilizing the interferometer path length, and does not need an external reference for demodulation using a lock-in amplifier. Instead, we time-resolve the interferometer vibrations (which will appear as phase modulation on our signals) by sampling with a high-speed digitizer at ~28 kHz. All three of the degenerate FWM signals, a time-resolved linear absorption signal (C-LO), and an autocorrelation signal between arms (A-B) are recorded synchronously at different radio frequencies. Simultaneously, we send a strongly red-detuned (1020 nm) reference laser through the interferometers and detect modulation beats on a separate detector. The reference laser samples the interferometer vibrations and position as a delay stage is scanned continuously. Because the reference beat note is phase sensitive, modulation caused by the interferometer vibrations manifests as relative Doppler shifts in the interferometer arms. The strongly red-detuned reference laser (relative to the oscillator at 796 nm) is an improvement over previous schemes that require the use of a degenerate optical reference because it avoids background population due to reference laser excitation, does not rely on stage encoders for interferometer position, and does not need to be tuned if the oscillator wavelength is adjusted. The effect of post-processing is to create a super-heterodyne optical receiver with feed-forward carrier phase recovery. Thus, the FWM signals cans be shifted to an arbitrary rotating optical frame using this reference. This experiment is appropriate for measuring long-lived coherences and recombination lifetimes limited only by the stage length.

3. Results

To demonstrate our technique, we resonantly excite a four-quantum-well sample with 100 fs pulses (at 796 nm). Our sample consists of four 8-nm AlGaAs/GaAs quantum wells grown on top of a semiconductor Bragg mirror. All three-degenerate wave-mixing signals (rephasing, non-rephasing, and two-quantum) are recorded simultaneously at their respective demodulation frequencies. The amplitude (Fig.1B) of a typical rephasing pulse sequence demonstrates three orders of magnitude of SNR with an excitation power of 13 μW (spot size ~9 μm). The spectrum has two diagonal peaks corresponding to a heavy-hole exciton and a weak trion which are coupled as indicated by the presence of off-diagonal cross-peaks. We attribute the presence of the weak trion to unintentional background doping in the growth process. The real part of the of the rephasing FWM signal (Fig. 1C) is purely absorptive, demonstrating the excellent ability to recover the carrier phase of the FWM signal.

4. Conclusion

In summary, we have demonstrated a technique that measures all four-wave mixing quantum pathways simultaneously in a *single* measurement. This experiment does not rely on wave-vector matching in the far field, can operate with diffraction-limited spatial resolution (*i.e.*, high numerical aperture objectives), and leverages heterodyne detection making it uniquely suited to optical experiments on single and few quantum emitters. Further, by using an optical reference laser recorded with better than 60-70 dB of signal-to-noise, our achievable phase stability and timing resolution is better than 20 attoseconds.

5. References

[1] P. Tian, D. Keusters, Y. Suzaki, W.S. Warren, "Femtosecond Phase-Coherent Two-Dimensional Spectroscopy," *Science* **300,** 1553–1555 (2003).

[2] P. F. Tekavec, G. A. Lott, A. H. Marcus, "Fluorescence-detected two-dimensional electronic coherence spectroscopy by acousto-optic phase modulation," *J. Chem. Phys.* **127,** 214307 (2007).

[3] M. Aeschlimann, et al. "Coherent Two-Dimensional Nanoscopy," *Science* **333,** 1723–1726 (2011).

[4] G. Nardin, T. M. Autry, K. L. Silverman, S. T. Cundiff, "Multidimensional coherent photocurrent spectroscopy of a semiconductor nanostructure," *Opt. Express* **21,** 28617–28627 (2013).

[5] V. Delmonte, et al. "Coherent coupling of individual quantum dots measured with phase-referenced two-dimensional spectroscopy: Photon echo versus double quantum coherence," *Phys. Rev. B* **96,** 41124 (2017).

[6] E. W. Martin, S. T. Cundiff, "Controlling Coherent Quantum Dot Interactions," *arXiv:1705.04730 [cond-mat]* (2017).

Autry, Travis; Moody, Galan; Fraser, James; McDonald, Corey; Mirin, Richard; Silverman, Kevin.
"Vibrational Interferometry Enables Single-Scan Acquisition of all {chi}^u(3)^ Multi-Dimensional Coherent Spectral Maps."
Paper presented at Conference on Lasers and Electro-Optics & Quantum Electronics and Laser Science Conference (CLEO/QELS),
San Jose, CA, United States. May 13, 2018 - May 18, 2018.

Developing a Programmable STEM Detector for the Scanning Electron Microscope

Benjamin W. Caplins[1], Jason D. Holm[1] and Robert R. Keller[1]

[1.] Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, CO, United States.

The scanning electron microscope (SEM) is traditionally used to analyze bulk samples, relying on signals generated by secondary and reflected/back-scattered electrons and photons to generate images. For relatively thin samples, forward-scattered electrons contain a wealth of information derived from the scattering volume. Diffraction within crystalline samples generates Bragg and Kikuchi scattered electrons which are sensitive to the crystallography and orientation of the material in addition to the presence of crystal defects. Scattering within amorphous samples leads to incoherent scattering relating to the mass thickness and/or atomic number of the material while coherent scattering can be related to the pairwise radial distribution function of the constituent atoms. Due to the reduced scattering volume of transmitted electrons relative to reflected electrons, the information derived from transmission measurements is generally obtained at a resolution equal-to or better-than bulk techniques. Scanning transmission electron microscopy (STEM) in the SEM (STEM-in-SEM) a promising technique to characterize nanoscale systems without the equipment overhead of a dedicated STEM.

One difficulty in making practical use of the forward-scattered electrons lies in the limitations of current STEM detectors. Broadly, STEM detectors can be categorized as either imaging detectors or integrating detectors. Imaging detectors give the user direct access to the sample's diffraction pattern at each beam position, but are generally slow (~1000 frames/second) and require (and permit) extensive off-line data analysis. On the other hand, integrating detectors can operate at high speeds (kHz to MHz) with excellent signal-to-noise, but cannot easily determine what features in the diffraction pattern are being measured, making image interpretation potentially difficult or even impossible in some situations.

Herein, we share progress made in our laboratory towards developing a new type of STEM detector that offers a hybrid approach. The basic detector design is shown in Figure 1a (see Ref. [1] for details). Transmitted electrons strike a scintillator screen placed directly beneath the sample, which serves to convert the electrons into photons. These photons are first imaged to a digital micromirror device (DMD) and then reimaged to a CMOS camera and/or an integrating photodetector (PMT). The DMD is a two-dimensional array (1024x768) of individually addressable mirrors which can be rapidly programmed to tilt up or down (towards the CMOS camera or PMT).

This programmable STEM (p-STEM) detector is currently used in two distinct modes. In `diffraction' mode, all of the mirrors are tilted towards the CMOS camera giving direct access to the full diffraction pattern from a sample (Figure 1b). In `imaging' mode, a user-defined subset of the mirrors is tilted towards the integrating photodetector which is synchronized with the microscope scan generator to create a real-space image. The DMD allows the user to easily switch between these modes during imaging sessions. A careful determination of the mapping functions between the scintillator, DMD, and CMOS camera allows the user to relate each mirror on the DMD to a scattering direction (Figure 1c).

Proof-of-principle data on a range of samples show diffraction contrast. In particular, we highlight the ability to generate contrast based on the crystallographic orientation of graphene. A dark-field image

Caplins, Benjamin; Keller, Robert; Holm, Jason.
"Developing a Programmable STEM Detector for the Scanning Electron Microscope."
Paper presented at Microscopy & Microanalysis 2018, Baltimore, MD, United States. August 5, 2018 - August 9, 2018.

based on integrating reflections of a particular graphene grain shows contrast completely absent in the secondary-electron and bright-field image of the same area (Figure 2a). A color-composite of several different dark-field images from a large field-of-view is shown in Figure 2b.

The presented data show that the p-STEM detector is a promising new detector design for STEM-in-SEM. Directly inspecting the diffraction pattern from a sample allows the user to orient their sample and select the appropriate electron beam parameters. The ability to quantitively define an arbitrary region of the diffraction pattern to integrate allows the user to specify dark-field conditions beyond simple circular or annular masks.

[1] B Jacobson *et al*, Proceedings of SPIE (2015) p. 93760K.

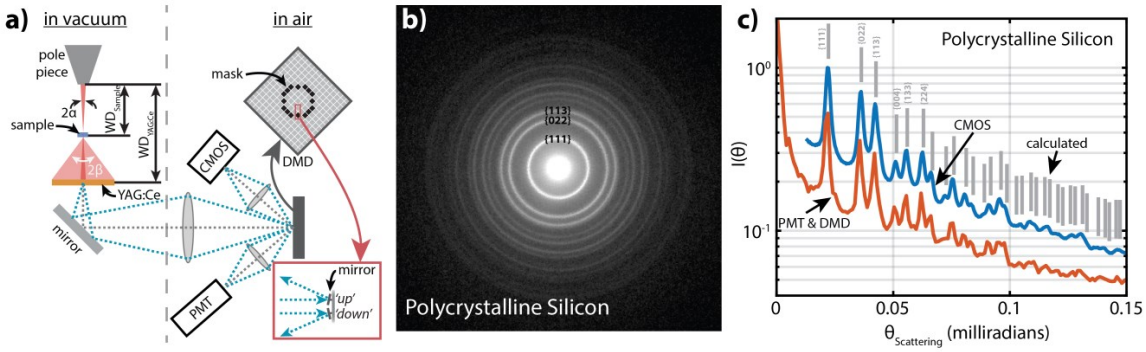


Figure 1. a) Electron detector schematic. The forward-scattered electrons are converted to photons on a scintillator screen which are optically imaged first to a DMD and then to a CMOS camera and/or integrating photodetector (PMT). b) A diffraction pattern from a polycrystalline silicon sample collected with all the DMD mirrors tilted towards the CMOS camera. c) Vertical bars: Calculated diffraction peak locations for silicon. Top curve: The annular integration of a polycrystalline silicon diffraction pattern collected on the CMOS camera. Bottom curve: Scattering data collected on the same sample by measuring the output of an integrating photodetector while programming a series of thin annular masks onto the DMD.

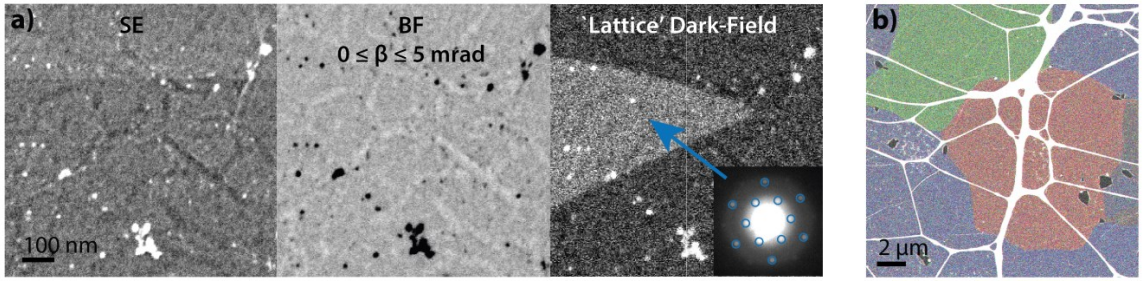


Figure 2. a) Secondary-electron, bright-field, and `lattice' dark-field images of monolayer graphene. The `lattice' dark field was collected with a mask on the DMD corresponding to the indicated regions of the diffraction pattern in the inset. b) An (RGB) color-composite image of three different `lattice' dark field images over a large field of view (ca. 19 µm). Individual grains of graphene have color contrast, while the support film is white.

Evaluation of Lateral and Depth Resolutions of Light Field Cameras

Soweon Yoon[1,3], Peter Bajcsy[1], Maritoni Litorja[2], and James J. Filliben[1]

[1.] Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA
[2.] Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA
[3.] Dakota Consulting Inc., Silver Spring, MD, USA

Light field cameras, also called plenoptic cameras, capture a field of light rays traveling in space, i.e., the intensity and direction of light rays. This is contrary to conventional still-picture cameras that acquire the aggregated intensity of incident light rays from all directions. A light field camera can be achieved by placing an array of microlenses between the imaging sensor and the main lens. Since a light field camera is essentially viewed as multiple cameras in a 2D array, it can function as a 3D camera based on multiple-view geometry. In crime scene investigations, 3D forensic evidence such as tire tread and shoe imprints in substances like mud or snow can often provide useful information to identify suspects and victims. This work focuses on evaluating lateral and depth resolutions of the light field cameras that are hand-held, easy to use, and affordable [1] to understand their suitability for forensic applications.

For the lateral resolution evaluation, the light field images of a resolution target plate with a 3 by 3 array of Siemens stars were collected to achieve a full factorial design with the following variables: (i) distance between the camera and the target plate (330 mm to 1030 mm with a step size of 50 mm), (ii) camera (two cameras from the same model), (iii) zoom level (level 1 for the field of view of 280 mm, and level 2 for that of 195 mm), and (iv) illumination (level 1 for ambient lighting, and level 2 with additional halogen lamps). Under each condition, five images were collected—local replicates—to observe the effect of random noise, for example, from the imaging sensor or the lamps. The entire data collection was repeated three times—global replicates—to observe any systematic errors coming from device setup and operation of the experiments. For the depth resolution evaluation, the images of a resolution target plate containing concentric circles with known spacing and thickness were collected. The resolution target plate was tilted at two known angles such that the depth of the points on the resolution target gradually changed from left to right. The rest of the experimental setup for the depth resolution evaluation was identical to the one for the lateral resolution evaluation. A total of 5400 light field images (1800 images for the lateral resolution evaluation and 3600 images for the depth resolution evaluation) were collected.

The lateral resolution was evaluated by a modulation transfer function (MTF) from the Siemens star, which measures the image contrast between black and white wedges in the Siemens star while the radius of a probe circle changes (see Fig. 1) [2]. The analysis results showed that (i) when a lower level of zoom was used, the lateral resolution tended to be stable regardless of distance between the camera and the target resolution plate, but when a higher level of zoom was used, it tended to decrease significantly with respect to distance, (ii) a higher zoom factor yielded a higher lateral resolution, (iii) two cameras of the same model did not show meaningful differences in lateral resolution, except for the difference in partial lens quality, (iv) the center region in the camera's field of view generally provided a better lateral resolution than the peripheral regions, and (v) the ambient lighting condition yielded a better lateral

Yoon, Sowon; Bajcsy, Peter; Litorja, Maritoni; Filliben, James.
"Evaluation of Lateral and Depth Resolutions of Light Field Cameras."
Paper presented at Microscopy & Microanalysis 2018, Baltimore, MD, United States. August 5, 2018 - August 9, 2018.

resolution than excessive illumination. While the results (i)-(iv) were largely expected, the result (v) was surprising. This result may imply that the increase in overall brightness of the image acquisition environment does not always improve the resolution power of the camera.

The depth resolution was evaluated by estimating the disparity between the points on the different concentric circles. The disparity is the movement of a scene point in the images when the camera viewpoint changes, and is represented as a slope in an epipolar-plane image (EPI) in the case of light field imaging [3]. It can be said that the depth resolution that is associated with the tilted angle of the resolution target plate is achieved if the disparity estimation along the horizontal line of the resolution plate shows a monotonic trend (see Fig. 2). The preliminary results showed that depth difference less than 1 mm was resolvable.

The datasets for the evaluation of lateral and depth resolutions of the light field cameras and the analysis results are available online with an interactive user interface [4,5].

References:

[1] Lytro, Inc., https://www.lytro.com/.
[2] ISO 12233, "Photography – Electronic still picture imaging – Resolution and spatial frequency responses", (2017).
[3] R Bolles *et al*, "Epipolar-plane image analysis: An approach to determining structure from motion", International Journal of Computer Vision **1** (1987), p. 7-55.
[4] https://isg.nist.gov/deepzoomweb/resources/lytroEvaluations/index.html
[5] Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

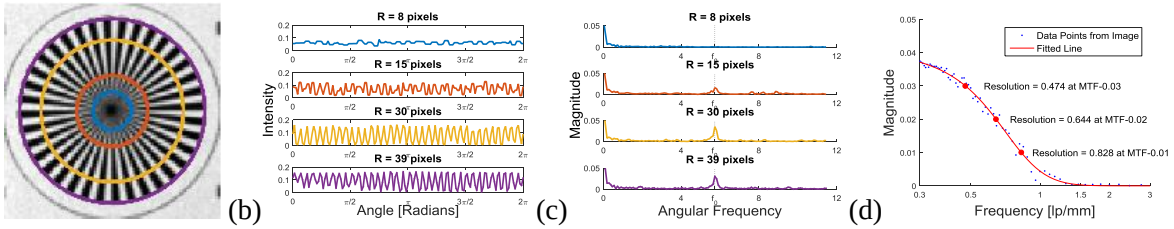


Figure 1. Measurement of lateral resolution. (a) Siemens star with four probe circles, (b) image intensity profile along each of the four circles, (c) Fourier transform of image intensity profiles, and (d) MTF.

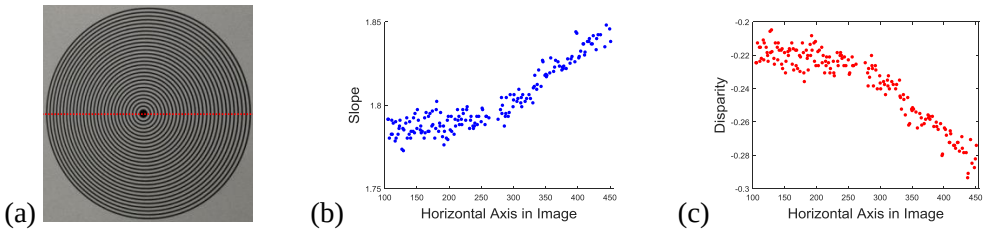


Figure 2. Measurement of depth resolution. (a) Concentric circles on a tilted plate, (b) slope estimation in EPI along the red line in (a), and (c) disparity estimation along the red line in (a).

Microscopic Identification of Micro-Organisms on Pre-Viking Swedish Hillfort Glass

Jamie L. Weaver[1], Carolyn I. Pearce[2], Bruce Arey[2], Michele Conroy[2], Edward P. Vicenzi[3], Rolf Sjoblom[4], Robert J. Koestler[3], Paula T. DePriest[3], Thomas F. Lam[3], David K. Peeler[2], John S. McCloy[5], and Albert A. Kruger[6]

[1.] National Institute of Standards and Technology, Gaithersburg, USA.
[2.] Pacific Northwest National Laboratory, Richland, USA.
[3.] Museum Conservation Institute, Smithsonian Institution, Suitland, USA.
[4.] Luleå University of Technology, Luleå, Sweden.
[5.] School of Mechanical and Materials Engineering, Washington State University, Pullman, USA.
[6.] Department of Energy, Office of River Protection, Richland, USA.

Broborg Hillfort is a vitrified hillfort located in the Husby-Långhundra parish in Uppland Sweden (Fig. 1A). The site dates to the Great Migration Period, around 740 ± 100 CE. [1]. Two types of glasses have been identified within the ramparts – one that is rich in Si and one that is rich in Fe [2]. Both glasses have been found on the outside surface of the walls, and have been exposed to atmosphere since their vitrification (Fig. 1B). Given these conditions, and their reported chemical compositions, the two glasses have been determined to be candidates for the study of the long-term alteration of silicate glasses in natural environments [3]. A unique aspect of the alteration research on these samples has been the use of microscopy to identify microbial activity on the surface of the glasses (Fig. 2 A-I). This presentation discusses how measurements of these species were made, and special techniques developed and used during sample preparation and data acquisition.

In this presented study, the surface of a specimen excavated from Broborg (Fig. 1C) was imaged in a FEI Helios NanoLab 660 (Hillsboro, OR) FIB-SEM with an Energy Dispersive Spectrometer (EDS) (EDAX Newark, NJ). An accelerating voltage of 5 keV for imaging, 10 or 20 keV for EDS (dependent chemistry/biochemistry of site), and a working distance of 4 mm were used in the analyses. Imaging was performed with an Everhart-Thornley secondary electron (SE) detector in field-free conditions, and through-the-lens detectors (TLD) for SE and back-scatter electron (BSE) imaging in immersion mode. Both glass and mineral sections of the sample were studied.

Certain commercial products are identified in this paper to specify the experimental procedures in adequate detail. This identification does not imply recommendation or endorsement by the authors or by the National Institute of Standards and Technology or Department of Energy, nor does it imply that the products identified are necessarily the best available for the purpose. Contributions of the Department of Energy and National Institute of Standards and Technology are not subject to copyright.

References:

[1] P. Kresten et al. Geochronometria, **22** (2003), p. 9.
[2] P. Kresten et al. Geologiska Föreningen i Stockholm Förhandlingar, **115** (1993), p. 13.
[3] R Sjöblom, H Ecke, & E Brännvall, Intl. J. Sust. Dev. Planning (2013).

Arey, Bruce; Vicenzi, Edward; Sjoblom, Rolf; Koestler, Robert; DePriest, Paula; Lam, Thomas; Peeler, David; McCloy, John; Kruger, Albert; Weaver, Jamie.
"Microscopic Identification of Micro-Organisms on Pre-Viking Swedish Hillfort Glass."
Paper presented at Microscopy & Microanalysis 2018, Baltimore, MD, United States. August 5, 2018 - August 9, 2018.

Figure 1. A) An arial view of Broborg. B) Excavation at Broborg, Fall 2017. C) Specimen of vitrified material analyzed in this study.

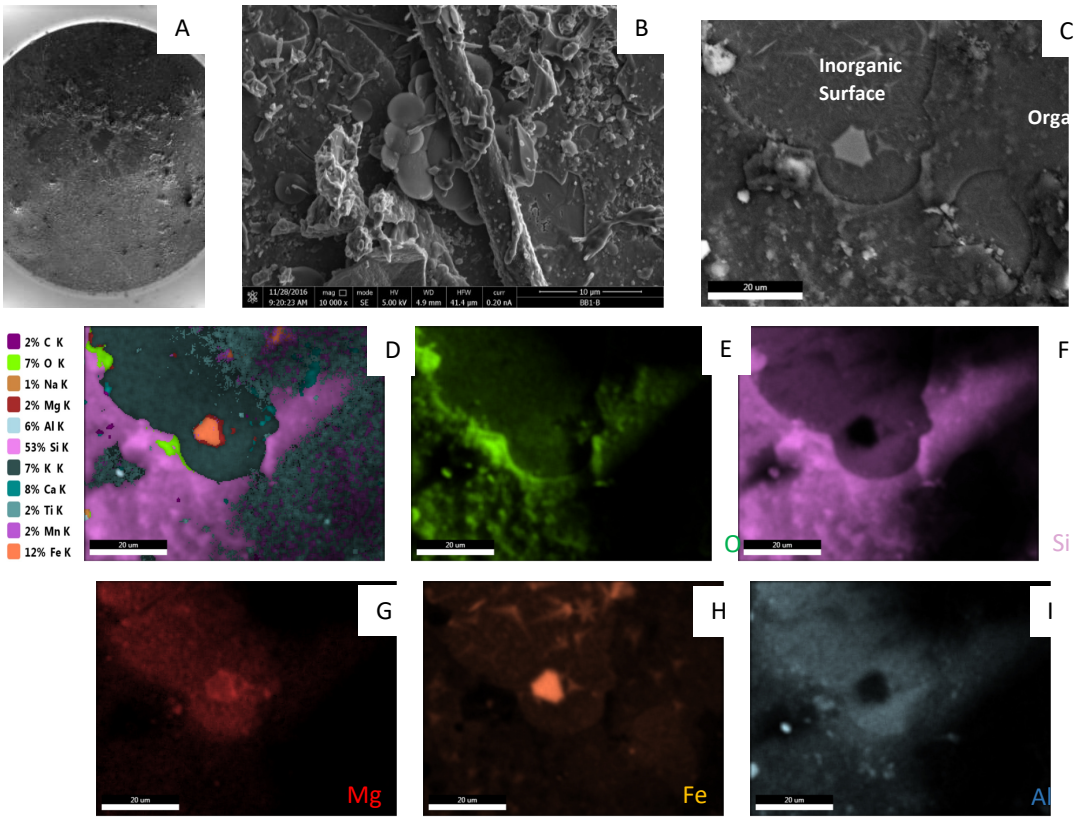


Figure 2. A) Low magnification image of glass sample with organics on the surface. B) High magnification image of glass surface with organics anpd microbes. C – I) SEM and EDS images of crystal in the high Fe glass matrix. D) is a composite image of SEM image and EDS maps for several elements measured for this sample area.

Arey, Bruce; Vicenzi, Edward; Sjoblom, Rolf; Koestler, Robert; DePriest, Paula; Lam, Thomas; Peeler, David; McCloy, John; Kruger, Albert; Weaver, Jamie.
"Microscopic Identification of Micro-Organisms on Pre-Viking Swedish Hillfort Glass."
Paper presented at Microscopy & Microanalysis 2018, Baltimore, MD, United States. August 5, 2018 - August 9, 2018.

Phase Retrieval Using Unitary 2-Designs

Shelby Kimmel* and Yi-Kai Liu*†
*Joint Center for Quantum Information and Computer Science (QuICS)
University of Maryland, College Park, MD, USA
†National Institute of Standards and Technology (NIST)
Gaithersburg, MD, USA

***Abstract*—We consider a variant of the phase retrieval problem, where vectors are replaced by unitary matrices, i.e., the unknown signal is a unitary matrix U, and the measurements consist of squared inner products $|\mathrm{tr}(C^\dagger U)|^2$ with unitary matrices C that are chosen by the observer. This problem has applications to quantum process tomography, when the unknown process is a unitary operation.**

We show that PhaseLift, a convex programming algorithm for phase retrieval, can be adapted to this matrix setting, using measurements that are sampled from *unitary 4- and 2-designs*. In the case of unitary 4-design measurements, we show that PhaseLift can reconstruct all unitary matrices, using a near-optimal number of measurements. This extends previous work on PhaseLift using spherical 4-designs.

In the case of unitary 2-design measurements, we show that PhaseLift still works pretty well on average: it recovers almost all signals, up to a constant additive error, using a near-optimal number of measurements. These 2-design measurements are convenient for quantum process tomography, as they can be implemented via randomized benchmarking techniques. This is the first positive result on PhaseLift using 2-designs.

I. Introduction

A. Phase Retrieval for Unitary Matrices

Phase retrieval is the problem of reconstructing an unknown vector $x \in \mathbb{C}^d$ from measurements of the form $|\langle a_i, x\rangle|^2$ (for $i = 1, \ldots, m$), where the $a_i \in \mathbb{C}^d$ are known vectors. This problem has been studied extensively in a number of contexts, including X-ray crystallography and optical imaging [1], [2].

In this paper we introduce a variant of the phase retrieval problem, where the vectors are replaced by *unitary matrices*: we want to reconstruct an unknown unitary matrix $U \in \mathbb{C}^{d\times d}$ from measurements of the form $|\mathrm{tr}(C_i^\dagger U)|^2$ (for $i = 1, \ldots, m$), where the $C_i \in \mathbb{C}^{d\times d}$ are known unitary matrices. (Here, $C_i^\dagger$ denotes the adjoint of the matrix C_i, and C_i^* denotes the complex conjugate.) This problem has applications in quantum process tomography, in the scenario where the experimenter wants to characterize an unknown unitary operation [3]–[5].

From a mathematical point of view, this problem can be seen to be a special case of phase retrieval, where the measurements have some additional algebraic structure: in addition to being vectors in $\mathbb{C}^{d^2}$, they are elements of the $d \times d$ unitary group $U(\mathbb{C}^{d\times d})$. This can be compared with previous work on phase retrieval, where the measurements were vectors in the unit sphere $S^{d-1} \subset \mathbb{C}^d$. In particular, a recent line of work has led to tractable algorithms for phase retrieval, for measurements that are *random vectors* in S^{d-1}, sampled from the Haar distribution, or from spherical 4-designs [6]–[10]. (We will define these terms more precisely in the following sections.)

The main contribution of this paper is to prove analogous results when the measurements are *random matrices* in $U(\mathbb{C}^{d\times d})$, sampled from the Haar distribution, or from unitary 4-designs. In addition, this paper proves weaker recovery guarantees when the measurements are sampled from spherical and unitary 2-designs. (These measurements are of interest because they are easier to implement in experiments, e.g., for quantum process tomography.) In the following sections, we will describe these results in more detail.

As a side note, while we have argued that it is natural to consider measurement matrices C_i that are unitary, one may ask whether it is still possible to recover *all* matrices U, and not just unitary ones? Unfortunately, the answer is no. For example, in the $d = 2$ case, consider the matrices

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } V = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{I.1}$$

These matrices cannot be distinguished using any measurement of the form $U \mapsto |\mathrm{tr}(C^\dagger U)|^2$, where $C \in \mathbb{C}^{2\times 2}$ is unitary. (To see this, note that any such C can be written in the form

$$C = \begin{pmatrix} \alpha & e^{i\varphi}\beta^* \\ \beta & -e^{i\varphi}\alpha^* \end{pmatrix}, \tag{I.2}$$

where $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$, and $\varphi \in \mathbb{R}$. Hence we have $|\mathrm{tr}(C^\dagger U)|^2 = |\alpha|^2 = |\mathrm{tr}(C^\dagger V)|^2$, i.e., the measurement results are the same for U and V.) Thus, the best we can hope for is to reconstruct some large *subset* of matrices in $\mathbb{C}^{d\times d}$, such as the set of all unitary matrices.

B. PhaseLift Algorithm

PhaseLift is a tractable algorithm for solving the phase retrieval problem, which works by "lifting" the quadratic problem in x to a linear problem in $xx^\dagger$ [11], [12]. We propose the following variant of PhaseLift, for phase retrieval of unitary matrices.

Suppose we want to recover an unknown unitary matrix $U \in \mathbb{C}^{d\times d}$. Our approach will be to solve for a Hermitian matrix $\Gamma \in \mathbb{C}^{d^2\times d^2}_{\mathrm{Herm}}$, in such a way that the solution has the form

$$\Gamma_{\mathrm{ideal}} = \mathrm{vec}(U)\mathrm{vec}(U)^\dagger, \tag{I.3}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

from which we can reconstruct U (up to a global phase factor). Here, vec denotes the map vec : $\mathbb{C}^{d\times d} \to \mathbb{C}^{d^2}$ that "flattens" a $d \times d$ matrix into a d^2-dimensional vector,

$$\mathrm{vec}(U) = (U_{1,1}, U_{1,2}, U_{1,3}, \ldots, U_{d,d-1}, U_{d,d})^T. \tag{I.4}$$

Note that this map preserves inner products: $\langle \mathrm{vec}(U), \mathrm{vec}(V)\rangle = \mathrm{tr}(U^\dagger V)$.

We can describe our quadratic measurements of U as follows. Let $C_1, \ldots, C_m \in \mathbb{C}^{d\times d}$ be the unitary measurement matrices introduced earlier. We use the same normalization convention as in previous work [9]: we work with re-scaled matrices $\sqrt{d}C_i$, which have roughly the same magnitude, in Frobenius or ℓ_2 norm, as Gaussian random matrices.[1]

We define the measurement operator $\mathcal{A} : \mathbb{C}^{d^2\times d^2}_{\text{Herm}} \to \mathbb{R}^m$ as follows:

$$\mathcal{A}(\Gamma) = \left[\mathrm{vec}(\sqrt{d}C_i)^\dagger \Gamma \mathrm{vec}(\sqrt{d}C_i)\right]_{i=1}^m. \tag{I.5}$$

Given the unknown matrix U, our measurement process returns a vector

$$\begin{aligned} y &= \mathcal{A}\big(\mathrm{vec}(U)\mathrm{vec}(U)^\dagger\big) + \varepsilon \\ &= \left[d|\mathrm{tr}(C_i^\dagger U)|^2\right]_{i=1}^m + \varepsilon, \end{aligned} \tag{I.6}$$

where $\varepsilon \in \mathbb{R}^m$ is an additive noise term.

Given the measurement results y, and an upper-bound $\eta \geq \|\varepsilon\|_2$ on the strength of the noise (measured using the ℓ_2 norm), we will then find Γ by solving a convex program:

$$\begin{aligned} \arg \min_{\Gamma \in \mathbb{C}^{d^2\times d^2}_{\text{Herm}}} & \mathrm{tr}(\Gamma) \text{ such that} \\ & \|\mathcal{A}(\Gamma) - y\|_2 \leq \eta, \\ & \Gamma \succeq 0, \\ & \mathrm{tr}_1(\Gamma) = (I/d)\,\mathrm{tr}(\Gamma), \\ & \mathrm{tr}_2(\Gamma) = (I/d)\,\mathrm{tr}(\Gamma). \end{aligned} \tag{I.7}$$

Here, $\mathrm{tr}_1(\Gamma) \in \mathbb{C}^{d\times d}$ and $\mathrm{tr}_2(\Gamma) \in \mathbb{C}^{d\times d}$ denote the partial traces over the first and second indices of Γ, defined as follows. We view $\Gamma \in \mathbb{C}^{d^2\times d^2}$ as a matrix whose entries $\Gamma_{(a,i),(b,j)}$ are indexed by $(a,i),(b,j) \in \{1,\ldots,d\}^2$. Then $\mathrm{tr}_1(\Gamma)$ and $\mathrm{tr}_2(\Gamma)$ are the matrices whose entries are defined by

$$\mathrm{tr}_1(\Gamma)_{i,j} = \sum_{a=1}^d \Gamma_{(a,i),(a,j)}, \quad \mathrm{tr}_2(\Gamma)_{a,b} = \sum_{i=1}^d \Gamma_{(a,i),(b,i)}. \tag{I.8}$$

The last three constraints in (I.7) have the effect of forcing Γ to be a linear combination of terms $\mathrm{vec}(V)\mathrm{vec}(V)^\dagger$ where the V are unitary. (This follows from some standard facts in quantum information theory. Let $|\Phi^+\rangle$ denote the maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |i\rangle \otimes |i\rangle \in \mathbb{C}^{d^2}$. Note that Γ is proportional to the Jamiolkowski state $J(\mathcal{E}) = (\mathcal{E}\otimes\mathcal{I})(|\Phi^+\rangle\langle\Phi^+|) \in \mathbb{C}^{d^2\times d^2}$ of some quantum process $\mathcal{E} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$. The last two constraints in (I.7) imply that $\mathrm{tr}_1(J(\mathcal{E})) = \mathrm{tr}_2(J(\mathcal{E})) = I/d$. This implies that $\mathcal{E}$ is unital and trace-preserving, which implies that $\mathcal{E}$ is an affine combination of unitary processes $\mathcal{V}$ [13]. Hence $J(\mathcal{E})$ is an affine combination of terms $J(\mathcal{V}) = \frac{1}{d}\mathrm{vec}(V)\mathrm{vec}(V)^\dagger$, where the V are unitary.)

Also, note that the desired solution $\Gamma_{\text{ideal}} = \mathrm{vec}(U)\mathrm{vec}(U)^\dagger$ satisfies these constraints, since we have $\mathrm{tr}_1(\Gamma_{\text{ideal}}) = (U^\dagger U)^T = I$, $\mathrm{tr}_2(\Gamma_{\text{ideal}}) = UU^\dagger = I$, and $\mathrm{tr}(\Gamma_{\text{ideal}}) = \mathrm{tr}(U^\dagger U) = d$.

C. Random Measurements and Unitary 4-Designs

We will consider the situation where the measurement matrices $C_1, \ldots, C_m$ are chosen independently at random from some probability distribution over the unitary group $U(\mathbb{C}^{d\times d})$. We will be interested in several choices for this distribution over $U(\mathbb{C}^{d\times d})$. One natural choice is the unitarily-invariant distribution, often called the *Haar distribution*, because this gives a well-defined notion of a "uniformly random" unitary matrix. However, for practical purposes, Haar-random measurements are often difficult to implement, when the dimension d is large.

This motivates us to consider *unitary t-designs*, which are distributions that have the same t'th-order moments as the Haar distribution. Formally, we say that a distribution $\mathcal{D}$ on $\mathbb{C}^{d\times d}$ is a unitary t-design if

$$\int_{\mathcal{D}} V^{\otimes t} \otimes (V^\dagger)^{\otimes t}\, dV = \int_{\text{Haar}} V^{\otimes t} \otimes (V^\dagger)^{\otimes t}\, dV. \tag{I.9}$$

(Here, $A \otimes B$ denotes the Kronecker product of two matrices A and B, and $A^{\otimes t}$ denotes the t-fold Kronecker product $A \otimes A \otimes \cdots \otimes A$.) Unitary t-designs have been studied in quantum information theory, where they are used to perform tasks such as quantum encryption, fidelity estimation and decoupling [14]–[18], [33]. For many purposes, it is sufficient to use unitary t-designs where t is much smaller than d, e.g., $t = 2$, $t = 4$, or $t = \mathrm{poly}(\log d)$. In these cases, there are explicit and computationally efficient constructions for these designs [19]–[25].

We will show that the PhaseLift algorithm in equation (I.7) succeeds in reconstructing the unknown matrix U, when the measurement matrices $C_1, \ldots, C_m$ are chosen at random from a *unitary 4-design*. In particular, we prove a recovery theorem that has a number of attractive features. First, the number of measurements m is close to optimal (up to a factor of $\log d$), since the unknown matrix U has $\Omega(d^2)$ degrees of freedom. Second, the failure probability is exponentially small in m, and the resulting recovery guarantee is uniform over all U. Third, the recovery is robust to noise, with an explicit bound on the error as a function of η and m.

Our proof builds on the work of [9], who showed an analogous result for PhaseLift when the measurements are random vectors sampled from a spherical 4-design. We use the same high-level approach, known as Mendelson's small ball method [28]–[30]. Our main technical innovation is the use of diagrammatic calculus and Weingarten functions [31]–[33],

[1] To justify this choice, we recall the definition of the Frobenius norm, $\|M\|_F = \mathrm{tr}(M^\dagger M)^{1/2} = (\sum_{ij}|M_{ij}|^2)^{1/2}$. For our re-scaled matrices $\sqrt{d}C_i$, we have $\|\sqrt{d}C_i\|_F^2 = \mathrm{tr}(\sqrt{d}C_i^\dagger C_i\sqrt{d}) = d^2$. For comparison, we note that a Gaussian random matrix $G \in \mathbb{C}^{d\times d}$, whose entries $G_{ij} \in \mathbb{C}$ are sampled independently from a complex Gaussian distribution with mean 0 and variance 1, has expected squared norm $\mathbb{E}[\|G\|_F^2] = d^2$.

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

in order to bound the 4th-moment tensor shown in equation (I.9).

Formally, we prove the following bound on the solution that is returned by the PhaseLift algorithm:

Theorem I.1. *There exists a numerical constant $c_5 > 0$ such that the following holds. Fix any $c_0 > 2c_5$. Consider the above scenario where m measurement matrices $C_1, \dots, C_m$ are chosen independently at random from a unitary 4-design $\hat{G}$ in $\mathbb{C}^{d\times d}$.*

Suppose that the number of measurements satisfies

$$m \geq \left(64(4!)^2 c_0\right)^2 \cdot d^2 \ln d. \tag{I.10}$$

Then with probability at least $1 - \exp\left(-2m\,(4(4!))^{-4}\right)$ (over the choice of the C_i), we have the following uniform recovery guarantee:

For any unitary matrix $U \in \mathbb{C}^{d\times d}$, it is the case that any solution Γ_{opt} to the convex program (I.7) with noisy measurements (I.6) must satisfy:

$$\|\Gamma_{opt} - vec(U)vec(U)^\dagger\|_F \leq \frac{128(4!)^2\eta}{\sqrt{m}}\left(1 + \frac{2c_5}{c_0 - 2c_5}\right). \tag{I.11}$$

We will present the proof in Section V.

D. Phase Retrieval Using Unitary 2-Designs

Next, we ask the question: how well does PhaseLift perform when the measurement matrices are sampled from a unitary 2-design, rather than a 4-design?

This question is motivated by a number of practical considerations. First, many experimental methods for quantum tomography make use of random Clifford operations [26], [27], which form a unitary 3-design, but not a 4-design [23]–[25]. Moreover, it is generally easier to construct unitary 2-designs, either using quantum circuits or group-theoretic methods [17], [19], [22]. The latter are particularly convenient for randomized benchmarking tomography, where the group structure is essential [48], [49].

We show that in this situation, PhaseLift still achieves *approximate* recovery of *almost all* unitary matrices. Here, the number of measurements m is still $O(d^2\text{poly}(\log d))$, which is close to optimal. "Approximate recovery" means that the algorithm recovers the desired solution $\text{vec}(U)\text{vec}(U)^\dagger$ up to an additive error of size $\delta\|\text{vec}(U)\text{vec}(U)^\dagger\|_F = \delta d$, where $\delta \ll 1$ is some constant that is independent of the dimension d. We show that this happens for all unitary matrices $U \in \mathbb{C}^{d\times d}$, except for a subset that is small with respect to Haar measure.

In addition, we prove an analogous result for recovery of *vectors* using PhaseLift, when the measurements are sampled from a *spherical* 2-design. Again, we can show approximate recovery of almost all vectors in $\mathbb{C}^d$, using $m = O(d\text{poly}(\log d))$ measurements.

These results give a clearer picture of the kinds of measurements that are sufficient for PhaseLift to succeed. In a sense, 2-designs are *almost* sufficient: while PhaseLift can sometimes fail using 2-design measurements, our results show that these failures occur very infrequently. (An explicit example of such a failure was previously shown in [7]: there exists a spherical 2-design $\mathcal{D}$ in $\mathbb{C}^d$, and there exist vectors $x \in \mathbb{C}^d$, such that phase retrieval of x requires $m = \Omega(d^2)$ measurements.)

Our proofs require a new ingredient, which is a notion of *non-spikiness* of the unknown vector or matrix that one is trying to recover. Informally, we say that the unknown vector is "non-spiky" if it has small inner product with all of the possible measurement vectors. (This is reminiscent of previous work on low-rank matrix completion [34]–[36].)

Our proofs proceed in two steps: first, we show that *almost all* vectors are non-spiky, and then we prove that *all* non-spiky vectors can be recovered (uniformly) using PhaseLift. In the second step, the non-spikiness property plays a crucial role in bounding certain 4th-moment quantities, where we can no longer rely on the 4th moments of the measurement vectors, because the measurements are now being sampled from a 2-design.

We now state our results formally, focusing on the recovery of *matrices*. (We will describe our results on the recovery of *vectors* in Section III.)

Let $\tilde{G}$ be a finite set of unitary matrices in $\mathbb{C}^{d\times d}$. We say that a unitary matrix $U \in \mathbb{C}^{d\times d}$ is *non-spiky* with respect to $\tilde{G}$ (with parameter $\beta \geq 0$) if the following holds:

$$|\text{tr}(C^\dagger U)|^2 \leq \beta, \quad \forall C \in \tilde{G}. \tag{I.12}$$

Generally speaking, we will say that U is non-spiky when $\beta \ll d$, e.g., $\beta \leq \text{poly}(\log d)$. Our first result shows that, when the set $\tilde{G}$ is not too large, almost all unitary matrices U are non-spiky with respect to $\tilde{G}$.

Proposition I.2. *Choose $U \in \mathbb{C}^{d\times d}$ to be a Haar-random unitary matrix. Then for any $t \geq 0$, with probability at least $1 - 4e^{-t}$ (over the choice of U), U is non-spiky with respect to $\tilde{G}$, with parameter*

$$\beta = \tfrac{9\pi^3}{2}(t + \ln|\tilde{G}|). \tag{I.13}$$

We will be interested in cases where the vectors in $\tilde{G}$ form a unitary 2-design in $\mathbb{C}^{d\times d}$. In these cases, the set $\tilde{G}$ can be relatively small, i.e., sub-exponential in d. As an example, let $d = 2^n$, and let $\tilde{G} \subset \mathbb{C}^{d\times d}$ be the set of Clifford operations on n qubits, so we have $|\tilde{G}| \leq 2^{2n^2+3n}$ [26], [27]. Then with high probability, U is non-spiky with respect to $\tilde{G}$, with parameter $\beta = O(\log^2 d)$.

We then define a non-spiky variant of the PhaseLift algorithm. This consists of the convex program (I.7), together with the following additional constraint, which forces the solution to be non-spiky:

$$0 \leq \text{vec}(C)^\dagger \Gamma \text{vec}(C) \leq \beta, \quad \forall C \in \tilde{G}. \tag{I.14}$$

We prove that the following recovery guarantee for this algorithm:

Theorem I.3. *There exists a numerical constant $c_5 > 0$ such that the following holds. Fix any $\delta > 0$ and $c_0 > 2c_5$. Consider the above scenario where m measurement matrices $C_1, \dots, C_m$ are chosen independently at random from a unitary 2-design $\tilde{G} \subset \mathbb{C}^{d\times d}$. Let $\beta \geq 0$, and define $\nu = \beta/\delta$.*

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

Suppose that the number of measurements satisfies

$$m \geq \left(8c_0\nu^2\right)^2 \cdot d^2 \ln d. \tag{I.15}$$

Then with probability at least $1 - \exp(-\frac{1}{128}m\nu^{-4})$ *(over the choice of the* C_i*), we have the following uniform recovery guarantee:*

For any unitary matrix $U \in \mathbb{C}^{d\times d}$ *that is non-spiky with respect to* $\tilde{G}$ *(with parameter* β*, in the sense of (I.12)), it is the case that any solution* Γ_{opt} *to the convex program in (I.7) and (I.14), with noisy measurements (I.6), must satisfy:*

$$\begin{aligned}&\|\Gamma_{opt} - vec(U)vec(U)^\dagger\|_F \\ &\leq \max\left\{\delta\|vec(U)vec(U)^\dagger\|_F,\ \tfrac{16\eta\nu^2}{\sqrt{m}}\left(1 + \tfrac{2c_5}{c_0 - 2c_5}\right)\right\}.\end{aligned} \tag{I.16}$$

Note that m, the number of measurements, again has almost-linear scaling with d^2, which is the number of degrees of freedom in the unknown matrix U. In addition, m depends on the dimensionless quantity $\nu = \beta/\delta$, where β measures the non-spikiness of the unknown matrix U, and δ controls the accuracy of the solution Γ_{opt}. In typical cases, we will have $\beta = O(\text{poly}(\log d))$, and we will choose δ to be constant, hence we have $\nu = O(\text{poly}(\log d))$.

We will present the proofs of these results for phase retrieval of *matrices* in Section IV, and for phase retrieval of *vectors* in Section III.

E. Application to Quantum Process Tomography

Finally, we describe an application of our results to quantum process tomography, in the case where the unknown process corresponds to a unitary operation $U \in \mathbb{C}^{d\times d}$. This is sufficient to describe the dynamics of a closed quantum system, e.g., time evolution generated by some Hamiltonian, or the action of unitary gates in a quantum computer, including coherent (but not stochastic) errors. There exist fast methods for learning unitary processes, which achieve a "quadratic speedup" over conventional tomography, using ideas from compressed sensing and matrix completion [3]–[5], [40]–[44].

Here, we describe an alternative way of achieving this quadratic speedup, using phase retrieval. Our approach via phase retrieval has a significant advantage: the measurements can be performed in a way that is robust against state preparation and measurement errors (SPAM errors), by using randomized benchmarking techniques [45]–[49]. We discuss this in Section VI.

F. Outlook

In this paper we have studied a variant of the phase retrieval problem that seeks to reconstruct *unitary matrices*, and we have proposed a variant of the PhaseLift algorithm that solves this problem. We have proved strong reconstruction guarantees when the measurements are sampled from unitary 4-designs, as well as weaker guarantees when the measurements are unitary and spherical 2-designs. This leads to novel methods for quantum process tomography, when the unknown process is a unitary operation.

We mention a few interesting open problems. One is to prove error bounds that depend on the ℓ_1 norm of the noise, rather than the ℓ_2 norm, as in [8]. Another problem is to extend our recovery method to handle processes with Kraus rank $r > 1$ (e.g., stochastic errors), perhaps using the techniques in [9].

A third problem is to prove tighter bounds when the measurements are random Clifford operations. Clifford operations are particularly convenient for quantum tomography, and they play an essential role in randomized benchmarking methods. They are known to be a unitary 3-design, but not a 4-design [23]–[25]. However, our numerical simulations in Section VI seem to indicate that random Clifford measurements perform *better* than their classification as a unitary 3-design would suggest. This is supported by several recent results showing that random Clifford operations are "close to" a unitary 4-design [50], [51], and that random vectors chosen from a Clifford orbit perform well for phase retrieval of *vectors* [10]. Can one prove a similar result for phase retrieval of *matrices*, using random Clifford operations?

Finally, there has been progress in solving phase retrieval problems using gradient descent algorithms, such as Wirtinger Flow [52]. Can these methods be adapted to our matrix setting?

II. Outline of the Paper

In the rest of this paper, we will present the proofs of our theorems, as well as further details on quantum process tomography. In Section III, we begin with our simplest result, on phase retrieval of *vectors*, using measurements sampled from spherical 2-designs. In Section IV, we extend this to phase retrieval of unitary *matrices*, using measurements sampled from unitary 2-designs. In Section V, we extend this further, to handle measurements sampled from unitary *4-designs*. Finally, in Section VI, we present further details and numerical simulations regarding quantum process tomography.

A. Notation

Let $\mathbb{C}^{d\times d}_{\text{Herm}}$ be the set of $d \times d$ complex Hermitian matrices. Let $\mathcal{L}(\mathbb{C}^{d\times d}, \mathbb{C}^{d\times d})$ be the set of linear maps from $\mathbb{C}^{d\times d}$ to $\mathbb{C}^{d\times d}$. We will use calligraphic letters to denote these maps, e.g., $\mathcal{U}, \mathcal{C}$. In general we will consider unitary maps, where $\mathcal{U} : \rho \to U\rho U^\dagger$ for a unitary U, and $\mathcal{C} : \rho \to C\rho C^\dagger$ for a unitary C. We will use $\mathcal{U}$ to represent the unknown map we want to recover, and $\mathcal{C}$ to represent the measurement maps we compare with $\mathcal{U}$. Thus sometimes $\mathcal{C}$ will represent an element of a unitary 2-design, and sometimes it will represent an element of a unitary 4-design.

For any matrix A, let $A^\dagger$ be its adjoint, let A^T be its transpose, and let A^* be its complex conjugate.

For any matrix A, with singular values $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_d(A) \geq 0$, let $\|A\| = \sigma_1(A)$ be the operator norm, let $\|A\|_F = \sqrt{\sum_i \sigma_i^2(A)}$ be the Frobenius norm, and let $\|A\|_* = \sum_i \sigma_i(A)$ be the nuclear or trace norm.

Because we use very similar approaches for the three cases of spherical 2-designs, unitary 2-designs, and unitary 4-designs, we will have similar notation in each section. In general, un-addorned notation (e.g. f, $\mathcal{A}$) will be used for

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

spherical 2-designs, notation with tildes (e.g. $\tilde{f}$, $\tilde{\mathcal{A}}$) will be used for unitary 2-designs, and notation with hats (e.g. $\hat{f}$, $\hat{\mathcal{A}}$) will be used for unitary 4-designs.

III. Phase Retrieval and Low-Rank Matrix Recovery Using Spherical 2-Designs

We first consider phase retrieval of *vectors*. In fact, we follow the approach of [9], and consider the more general problem of low-rank matrix recovery. Whereas the authors of [9] showed an exact recovery result for measurements that are sampled from a spherical 4-design, we show an approximate, average-case recovery result for measurements that are chosen from a spherical 2-design.

We want to reconstruct an unknown matrix $X \in \mathbb{C}^{d\times d}_{\text{Herm}}$, having rank at most r, from quadratic measurements of the form

$$y_i = w_i^\dagger X w_i + \varepsilon_i, \qquad i = 1, 2, \dots, m, \tag{III.1}$$

where the measurement vectors $w_i \in \mathbb{C}^d$ are known and are sampled independently at random from a spherical 2-design, and the $\varepsilon_i \in \mathbb{R}$ are unknown contributions due to additive noise.

This problem has been studied previously, particularly in the rank-1 case (setting $r = 1$), where it corresponds to phase retrieval [7], [9]. Roughly speaking, it is known that spherical 4-designs are sufficient to recover X efficiently [9], whereas there exists a spherical 2-design that can not recover X efficiently [7]. More precisely, X can be recovered (uniformly), via convex relaxation, from $m = O(rd \operatorname{poly}\log d)$ measurements chosen from any spherical 4-design; while on the other hand, X cannot be recovered, by any method, from fewer than $m = \Omega(d^2)$ measurements chosen from a particular spherical 2-design.

Here, we show that 2-designs are sufficient to recover *a large subset* of all the rank-r matrices in $\mathbb{C}^{d\times d}_{\text{Herm}}$. More precisely, we show that 2-designs achieve efficient *approximate* recovery of all low-rank matrices X that are *non-spiky* with respect to the measurement vectors (we will define this more precisely below). This implies that 2-designs are sufficient to recover generic (random) low-rank matrices X, since these matrices satisfy the non-spikiness requirement with probability close to 1. Here, "efficient approximate recovery" means recovery up to an arbitrarily small constant error, using $m = O(rd \operatorname{poly}\log d)$ measurements, by solving a convex relaxation. This is reminiscent of results on non-spiky low-rank matrix completion [34]–[36].

A. Non-spikiness condition

Let G be a finite set of vectors in $\mathbb{C}^d$, each of length 1. We say that X is *non-spiky* with respect to G (with parameter $\beta \geq 0$) if the following holds:

$$|w^\dagger X w| \leq \frac{\beta}{d}\|X\|_*, \quad \forall w \in G. \tag{III.2}$$

Here, $\|X\|_* = \operatorname{tr}(|X|)$ denotes the nuclear norm. Generally speaking, we will say that X is non-spiky when the parameter β is much smaller than d, e.g., of size $\operatorname{poly}(\log d)$.

We now show that, when the set G is not too large, almost all rank-r matrices $X \in \mathbb{C}^{d\times d}_{\text{Herm}}$ are non-spiky with respect to G.

Proposition III.1. *Fix some rank-r matrix $W \in \mathbb{C}^{d\times d}_{\text{Herm}}$. Construct a random matrix X by setting $X = UWU^\dagger$, where $U \in \mathbb{C}^{d\times d}$ is a Haar-random unitary operator.*

Then for any $t \geq 0$, with probability at least $1 - 2e^{-t}$ (over the choice of U), X is non-spiky with respect to G, with parameter

$$\beta = 9\pi^3(t + \ln|G| + \ln r). \tag{III.3}$$

Proof: Let S^{d-1} denote the unit sphere in $\mathbb{C}^d$. Fix any vector $w \in S^{d-1}$, and let x be a uniformly random vector in S^{d-1}. Using Levy's lemma [53], [54], we have that

$$\Pr[|w^\dagger x| \geq \varepsilon] \leq 2\exp\left(-\tfrac{d\varepsilon^2}{9\pi^3}\right). \tag{III.4}$$

Setting $\varepsilon = \sqrt{\frac{9\pi^3}{d}(t + \ln|G| + \ln r)}$, we get that

$$\Pr[|w^\dagger x| \geq \varepsilon] \leq 2e^{-t}|G|^{-1}r^{-1}. \tag{III.5}$$

Now recall that $X = UWU^\dagger$, and write the spectral decomposition of W as $W = \sum_{i=1}^r \lambda_i v_i v_i^\dagger$. Then for any $w \in G$, we can write $w^\dagger X w$ as follows:

$$w^\dagger X w = w^\dagger UWU^\dagger w = \sum_{i=1}^r \lambda_i |w^\dagger U v_i|^2. \tag{III.6}$$

Each vector Uv_i is uniformly random in S^{d-1}, hence it obeys the bound in (III.5). Using the union bound over all $w \in G$ and all $v_1, \dots, v_r$, we conclude that:

$$|w^\dagger U v_i| \leq \varepsilon, \quad \forall w \in G,\ \forall i = 1, \dots, r, \tag{III.7}$$

with failure probability at most $2e^{-t}$. Combining this with (III.6) proves the claim. □

We will be interested in cases where the vectors in G form a spherical 2-design in $\mathbb{C}^d$. In these cases, the set G can be relatively small, i.e., sub-exponential in d. As an example, let $d = 2^n$, and let G be the set of stabilizer states on n qubits, so we have $|G| \leq 4^{n^2}$ [26]. Then with high probability, X is non-spiky with respect to G, with parameter $\beta = O(\log^2 d)$.

B. Convex relaxation (PhaseLift)

We now consider measurement vectors $w_1, \dots, w_m$ that are sampled independently from a spherical 2-design $G \subset \mathbb{C}^d$. Using these measurements, we will seek to reconstruct those matrices $X \in \mathbb{C}^{d\times d}_{\text{Herm}}$ that are low-rank, and non-spiky with respect to G.

We assume we are given a bound on the nuclear norm of X, say (without loss of generality):

$$\|X\|_* \leq 1. \tag{III.8}$$

As was done in [9], we will rescale the vectors w_i to be more like Gaussian random vectors, i.e., we will work with the vectors

$$a_i = [(d+1)d]^{1/4} w_i, \quad i = 1, \dots, m. \tag{III.9}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

We also let A be the renormalized version of the spherical 2-design G,

$$A = \{[(d+1)d]^{1/4} w \mid w \in G\}. \tag{III.10}$$

Then X will satisfy the following the non-spikiness conditions:

$$|a^\dagger X a| \le \beta\sqrt{1+\tfrac{1}{d}}, \quad \forall a \in A. \tag{III.11}$$

We define the sampling operator $\mathcal{A} : \mathbb{C}^{d\times d}_{\text{Herm}} \to \mathbb{R}^m$,

$$\mathcal{A}(Z) = \left(a_i^\dagger Z a_i\right)_{i=1}^m. \tag{III.12}$$

Using this notation, the measurement process returns a vector

$$y = \mathcal{A}(X) + \varepsilon, \quad \|\varepsilon\|_2 \le \eta, \tag{III.13}$$

where we assume we know an upper-bound η on the size of the noise term.

We will solve the following convex relaxation, which is essentially the PhaseLift convex program, augmented with non-spikiness constraints:

$$\begin{aligned} &\arg\min_{Z\in\mathbb{C}^{d\times d}_{\text{Herm}}} \|Z\|_* \text{ such that} \\ &\|\mathcal{A}(Z) - y\|_2 \le \eta, \\ &|a^\dagger Z a| \le \beta\sqrt{1+\tfrac{1}{d}}, \quad \forall a \in A. \end{aligned} \tag{III.14}$$

C. Approximate recovery of non-spiky low-rank matrices

We show the following uniform recovery guarantee for the convex relaxation shown in equation (III.14):

Theorem III.2. *Fix any $\delta > 0$ and $C_0 > 13$. Consider the above scenario where m measurement vectors $a_1, \dots, a_m$ are chosen independently at random from a (renormalized) spherical 2-design $A \subset \mathbb{C}^d$. Let $1 \le r \le d$ and $\beta \ge 0$, and define $\nu = 2\beta/\delta$.*

Suppose that the number of measurements satisfies

$$m \ge \left(8C_0\nu^2(1+\tfrac{1}{d})\right)^2 \cdot rd\log(2d). \tag{III.15}$$

Then with probability at least

$$1 - \exp\left(-\tfrac{1}{128}\, m\left(\nu^2(1+\tfrac{1}{d})\right)^{-2}\right)$$

(over the choice of the a_i), we have the following uniform recovery guarantee:

For any rank-r matrix $X_{true} \in \mathbb{C}^{d\times d}_{Herm}$ that has nuclear norm $\|X_{true}\|_* \le 1$, and is non-spiky with respect to A (with parameter β, in the sense of (III.11)), it is the case that any solution X_{opt} to the convex program (III.14) with noisy measurements (III.13) must satisfy:

$$\begin{aligned} &\|X_{opt} - X_{true}\|_F \\ &\quad \le \max\left\{\delta,\ \tfrac{16\eta\nu^2}{\sqrt{m}}(1+\tfrac{1}{d})(1+\tfrac{13}{C_0-13})\right\}. \end{aligned} \tag{III.16}$$

This error bound has similar scaling to the one in [9]: the number of measurements m is only slightly larger than the number of degrees of freedom $O(rd)$, and the error $X_{\text{opt}} - X_{\text{true}}$ decreases like $\eta/\sqrt{m}$, until it reaches the limit δ. However, now both of these bounds also involve the dimensionless quantity ν, which in turn depends on the non-spikiness β of the original matrix X_{true}, as well as the desired accuracy δ of the reconstructed matrix X_{opt}. In some sense, this is the price that one pays when one uses a weaker measurement ensemble, such as a spherical 2-design. It is an interesting question whether one can improve these bounds, to have a better dependence on ν.

The proof follows the same strategy used in [9] to show low-rank matrix recovery using spherical 4-design measurements, by means of Mendelson's small-ball method [30]. The key difference is that our measurements are spherical 2-designs only, so we do not have control over their fourth moments. Instead, we use the non-spikiness properties of X_{true} and X_{opt} to bound the fourth moments that appear in the proof. This allows us to show approximate recovery of X_{true}, up to an arbitrarily small constant error δ.

We will present this proof in several steps.

D. Approximate recovery via modified descent cone

We begin by defining a modified version of the descent cone used in [9], [30]. Let $f : \mathbb{R}^d \to \mathbb{R}\cup\{\infty\}$ be a proper convex function, and let $x_{\text{true}} \in \mathbb{R}^d$ and $\delta \ge 0$. Then we define the modified descent cone $D'(f, x_{\text{true}}, \delta)$ as follows:

$$\begin{aligned} D'(f, x_{\text{true}}, \delta) = \{y \in \mathbb{R}^d \mid \exists \tau > 0 \text{ such that} \\ f(x_{\text{true}} + \tau y) \le f(x_{\text{true}}), \\ \|\tau y\|_2 \ge \delta\}. \end{aligned} \tag{III.17}$$

This is the set of all directions, originating at the point x_{true}, that cause the value of f to decrease, when one takes a step of size at least δ. Note that this is a cone, but not necessarily a convex cone.

We use this to state an approximate recovery bound, analogous to Prop. 7 in [9] and Prop. 2.6 in [30]. Let y be a noisy linear measurement of x_{true}, given by $y = \Phi x_{\text{true}} + \varepsilon$, where $\Phi \in \mathbb{R}^{m\times d}$ and $\|\varepsilon\|_2 \le \eta$. Then let x_{opt} be a solution of the convex program

$$\arg\min_{z\in\mathbb{R}^d} f(z) \text{ such that } \|\Phi z - y\|_2 \le \eta. \tag{III.18}$$

Then we have the following recovery bound:

$$\|x_{\text{opt}} - x_{\text{true}}\|_2 \le \max\left\{\delta,\ \frac{2\eta}{\lambda_{\min}(\Phi; D'(f, x_{\text{true}}, \delta))}\right\}, \tag{III.19}$$

where $\lambda_{\min}$ is the conic minimum singular value,

$$\begin{aligned} &\lambda_{\min}(\Phi; D'(f, x_{\text{true}}, \delta)) \\ &\quad = \inf\{\|\Phi u\|_2 \mid u \in S^{d-1} \cap D'(f, x_{\text{true}}, \delta)\}. \end{aligned} \tag{III.20}$$

This is proved easily, using the same argument as in [9], [30].

We now re-state this bound for the setup in Theorem III.2. Here the function $f : \mathbb{C}^{d\times d}_{\text{Herm}} \to \mathbb{R}\cup\{\infty\}$ is given by

$$f(Z) = \begin{cases} \|Z\|_* & \text{if } z \text{ is non-spiky in} \\ & \text{the sense of (III.11),} \\ \infty & \text{otherwise,} \end{cases} \tag{III.21}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

and we use

$$D(f, X_{\text{true}}, \delta) = \{Y \in \mathbb{C}^{d\times d}_{\text{Herm}} \mid \exists \tau > 0 \text{ such that } f(X_{\text{true}} + \tau Y) \le f(X_{\text{true}}), \ \|\tau Y\|_F \ge \delta\}. \quad \text{(III.22)}$$

Our recovery bound is:

$$\|X_{\text{opt}} - X_{\text{true}}\|_F \le \max\left\{\delta, \ \frac{2\eta}{\lambda_{\min}(\mathcal{A}; D(f, X_{\text{true}}, \delta))}\right\}, \quad \text{(III.23)}$$

and we want to lower-bound the quantity $\lambda_{\min}$:

$$\lambda_{\min}(\mathcal{A}; D(f, X_{\text{true}}, \delta)) = \inf_{Y \in E(X_{\text{true}})} \left[\sum_{i=1}^m |\text{tr}(a_i a_i^\dagger Y)|^2\right]^{1/2},$$

where we define

$$E(X_{\text{true}}) = \{Y \in D(f, X_{\text{true}}, \delta) \mid \|Y\|_F = 1\}. \quad \text{(III.24)}$$

E. Mendelson's small-ball method

In order to lower-bound $\lambda_{\min}$, we will use Mendelson's small-ball method, following the steps described in [9] (see also [30]). We will prove a uniform lower-bound that holds for all X_{true} simultaneously, i.e., we will lower-bound $\inf_{X_{\text{true}}} \lambda_{\min}$. We define

$$E = \bigcup_{X_{\text{true}}} E(X_{\text{true}}), \quad \text{(III.25)}$$

where the union runs over all $X_{\text{true}} \in \mathbb{C}^{d\times d}_{\text{Herm}}$ that have rank at most r and satisfy the non- spikiness conditions (III.11). We then take the infimum over all $Y \in E$.

Using Mendelson's small-ball method (Theorem 8 in [9]), we have that for any $\xi > 0$ and $t \ge 0$, with probability at least $1 - e^{-2t^2}$,

$$\inf_{Y\in E}\left[\sum_{i=1}^m |\text{tr}(a_i a_i^\dagger Y)|^2\right]^{1/2} \ge \xi\sqrt{m} Q_{2\xi}(E) - 2W_m(E) - \xi t, \quad \text{(III.26)}$$

where we define

$$Q_\xi(E) = \inf_{Y\in E}\left\{\Pr_{a\sim\mathcal{A}}[|\text{tr}(aa^\dagger Y)| \ge \xi]\right\}, \quad \text{(III.27)}$$

$$W_m(E) = \mathop{\mathbb{E}}_{\substack{\epsilon_i\sim\pm 1\\ a_i\sim\mathcal{A}}}\left[\sup_{Y\in E} \text{tr}(HY)\right], \quad H = \frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i a_i a_i^\dagger, \quad \text{(III.28)}$$

and $\varepsilon_1, \dots, \varepsilon_m$ are Rademacher random variables.

F. Lower-bounding $Q_\xi(E)$

We will lower-bound $Q_\xi(E)$ as follows. Fix some $Y \in E$. We know that Y is in $E(X_{\text{true}})$, for some $X_{\text{true}} \in \mathbb{C}^{d\times d}_{\text{Herm}}$ that has rank at most r and is non-spiky in the sense of (III.11). Hence we know that $\|Y\|_F = 1$, and there exists some $\tau > 0$ such that $X_{\text{true}} + \tau Y$ is non-spiky in the sense of (III.11), and we have

$$\|X_{\text{true}} + \tau Y\|_* \le \|X_{\text{true}}\|_*, \quad \|\tau Y\|_F \ge \delta. \quad \text{(III.29)}$$

Note that we have $\|\tau Y\|_F = \tau \ge \delta$.

We will lower-bound $\Pr[|\text{tr}(aa^\dagger Y)| \ge \xi]$, for any $\xi \in [0, 1]$, using the Paley-Zygmund inequality, and appropriate bounds on the second and fourth moments of $\text{tr}(aa^\dagger Y)$. Let us define a random variable

$$S = |\text{tr}(aa^\dagger \tau Y)|. \quad \text{(III.30)}$$

We can lower-bound $\mathbb{E}(S^2)$, using the same calculation as in Prop. 12 of [9]:

$$\mathbb{E}(S^2) = \text{tr}(\tau Y)^2 + \|\tau Y\|_F^2 \ge 0 + \tau^2. \quad \text{(III.31)}$$

To handle $\mathbb{E}(S^4)$, we need to use a different argument from the one in [9], since our a is sampled from a spherical 2-design, not a 4-design. Our solution is to upper-bound S, using the non-spikiness properties of X_{true} and $X_{\text{true}} + \tau Y$:

$$S = \left|-\text{tr}(aa^\dagger X_{\text{true}}) + \text{tr}(aa^\dagger(X_{\text{true}} + \tau Y))\right| \le (2\beta)\sqrt{1 + \tfrac{1}{d}}. \quad \text{(III.32)}$$

This implies an upper-bound on $\mathbb{E}(S^4)$:

$$\mathbb{E}(S^4) \le (2\beta)^2(1 + \tfrac{1}{d})\mathbb{E}(S^2). \quad \text{(III.33)}$$

Putting it all together, we have:

$$\begin{aligned}\Pr_{a\sim\mathcal{A}}[|\text{tr}(aa^\dagger Y)| \ge \xi] &= \Pr[S^2 \ge \tau^2\xi^2] \\ &\ge \Pr[S^2 \ge \xi^2\mathbb{E}(S^2)] \\ &\ge (1-\xi^2)^2\frac{\mathbb{E}(S^2)^2}{\mathbb{E}(S^4)} \\ &\ge (1-\xi^2)^2\frac{\tau^2}{(2\beta)^2(1+\frac{1}{d})}.\end{aligned} \quad \text{(III.34)}$$

Finally, using the fact that $\tau \ge \delta$, we have that

$$Q_{1/2}(E) \ge (1-\xi^2)^2\frac{\delta^2}{(2\beta)^2(1+\frac{1}{d})}. \quad \text{(III.35)}$$

G. Upper-bounding $W_m(E)$

Next, we will upper-bound $W_m(E)$, using essentially the same argument as in [9]. Recall that the argument in [9] showed an upper-bound on $W_m(F)$, where F was a slightly different set than our E, and the a_i were sampled from a spherical 4-design rather than a 2-design. The argument used the following steps:

$$\begin{aligned}W_m(F) &= \mathbb{E}\sup_{Y\in F}\text{tr}(HY) \\ &\le \mathbb{E}\, 2\sqrt{r}\|H\| \\ &\le 2\sqrt{r}\cdot(3.1049)\sqrt{d\log(2d)},\end{aligned} \quad \text{(III.36)}$$

using Lemma 10 and Prop. 13 in [9], and provided that $m \ge 2d\log d$.

The first step still works to upper-bound $W_m(E)$, because $E \subset F$. To see this, recall that the set F was defined as follows:

$$F = \bigcup_{X_{\text{true}}} F(X_{\text{true}}), \quad \text{(III.37)}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

where the union was over all $X_{\text{true}} \in \mathbb{C}^{d\times d}_{\text{Herm}}$ with rank at most r, and

$$F(X_{\text{true}}) = \{Y \in D(\|\cdot\|_*, X_{\text{true}}, 0) \mid \|Y\|_F = 1\}. \quad \text{(III.38)}$$

This can be compared with our definition of the set E in (III.25) and (III.24).

The second step still works for our choice of the a_i, because Prop. 13 in [9] does not use the full power of the spherical 4-design. In fact it only requires that the a_i are sampled from a spherical 1-design (because the proof relies mainly on the Rademacher random variables ε_i). Thus we conclude that

$$W_m(E) \le 2\sqrt{r}\cdot(3.1049)\sqrt{d\log(2d)}. \quad \text{(III.39)}$$

H. Final result

Combining equations (III.24), (III.26), (III.35) and (III.39), and setting $\xi = \frac{1}{4}$, $\nu = 2\beta/\delta$ and $t = \frac{1}{16}\sqrt{m}\big(\nu^2(1+\frac{1}{d})\big)^{-1}$, we get that:

$$\begin{aligned}
&\inf_{X_{\text{true}}} \lambda_{\min}(\mathcal{A}; D(f, X_{\text{true}}, \delta)) \\
&\quad \ge \tfrac{1}{4}\sqrt{m}\,\frac{9}{16\nu^2(1+\frac{1}{d})} \\
&\qquad - (12.44)\sqrt{rd\log(2d)} - \tfrac{1}{4}t \\
&\quad = \tfrac{1}{8}\sqrt{m}\,\frac{1}{\nu^2(1+\frac{1}{d})} \\
&\qquad - (12.44)\sqrt{rd\log(2d)}.
\end{aligned} \quad \text{(III.40)}$$

Now we set $m \ge \big(8C_0\nu^2(1+\frac{1}{d})\big)^2 \cdot rd\log(2d)$, which implies

$$\frac{\sqrt{m}}{8C_0\nu^2(1+\frac{1}{d})} \ge \sqrt{rd\log(2d)}, \quad \text{(III.41)}$$

hence

$$\begin{aligned}
&\inf_{X_{\text{true}}} \lambda_{\min}(\mathcal{A}; D(f, X_{\text{true}}, \delta)) \\
&\qquad \ge \tfrac{1}{8}\sqrt{m}\,\frac{1}{\nu^2(1+\frac{1}{d})}\,\frac{C_0-13}{C_0}.
\end{aligned} \quad \text{(III.42)}$$

This can be plugged into our approximate recovery bound (III.23). This finishes the proof of Theorem III.2. □

IV. Phase Retrieval Using Unitary 2-Designs

Next, we extend our results from *vectors* to unitary *matrices*. First, we consider phase retrieval using measurements that are sampled from unitary 2-designs, as described in Section I-D.

A. Non-spikiness

First, let $\tilde{G} \subset \mathbb{C}^{d\times d}$ be a unitary 2-design. we prove Proposition I.2, showing that almost all unitary matrices are non-spiky with respect to $\tilde{G}$:

Proof: Fix any $C \in \tilde{G}$. Using Levy's lemma [56], and noting that the function $U \mapsto \mathrm{tr}(U^\dagger C)$ has Lipshitz coefficient $\eta \le \|C\|_F = \sqrt{d}$, we have that

$$\Pr[|\mathrm{tr}(U^\dagger C)| \ge \varepsilon] \le 4\exp\big(-\tfrac{2}{9\pi^3}\varepsilon^2\big). \quad \text{(IV.1)}$$

Setting $\varepsilon = \sqrt{\frac{9\pi^3}{2}(t + \ln|\tilde{G}|)}$, we get that

$$\Pr[|\mathrm{tr}(U^\dagger C)| \ge \varepsilon] \le 4e^{-t}|\tilde{G}|^{-1}. \quad \text{(IV.2)}$$

Using the union bound over all $C \in \tilde{G}$, we conclude that:

$$|\mathrm{tr}(U^\dagger C)| \le \varepsilon, \quad \forall C \in \tilde{G}, \quad \text{(IV.3)}$$

with failure probability at most $4e^{-t}$. This proves the claim. □

B. Approximate recovery of non-spiky unitary matrices

Next, let $\tilde{\mathcal{A}}$ denote the measurement operator defined in equation (I.5), where the measurement matrices $C_1, \dots, C_m$ are chosen independently at random from the unitary 2-design $\tilde{G}$.

We now prove Theorem I.3, showing that all non-spiky unitary matrices can be recovered approximately via PhaseLift, given these measurements. We will present this proof in several steps, following the same strategy as in the previous section.

We will use the following notation. For any unitary matrix $U \in \mathbb{C}^{d\times d}$, let $\mathcal{U} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ be the quantum process whose action is given by $\mathcal{U} : \rho \mapsto U\rho U^\dagger$. Let $J(\mathcal{U})$ be the corresponding Jamiolkowski state, as defined in equation (VI.3), which can be written as

$$J(\mathcal{U}) = \frac{1}{d}\mathrm{vec}(U)\mathrm{vec}(U)^\dagger \in \mathbb{C}^{d^2\times d^2}. \quad \text{(IV.4)}$$

Using this notation, we can write the desired solution of the PhaseLift convex program as $J(\mathcal{U})d$, and we can write the measurement operator $\tilde{\mathcal{A}}$ as

$$\tilde{\mathcal{A}}(\Gamma) = \Big[\mathrm{tr}(\Gamma J(\mathcal{C}_i)d^2)\Big]_{i=1}^m. \quad \text{(IV.5)}$$

We want to recover the solution $J(\mathcal{U})d$, up to an additive error of size $\delta\|J(\mathcal{U})d\|_F = \delta d$.

C. Modified descent cone

We define the function $\tilde{f} : \mathbb{C}^{d^2\times d^2}_{\text{Herm}} \to \mathbb{R}\cup\{\infty\}$ as follows:

$$\tilde{f}(\Gamma) = \begin{cases} \mathrm{tr}(\Gamma) & \text{if } \Gamma \text{ is non-spiky in the sense} \\ & \text{of (I.14), and also satisfies the} \\ & \text{last three constraints in (I.7),} \\ \infty & \text{otherwise.} \end{cases} \quad \text{(IV.6)}$$

Our recovery bound is:

$$\begin{aligned}
&\|\Gamma_{\text{opt}} - J(\mathcal{U})d\|_F \\
&\qquad \le \max\left\{\delta d,\ \frac{2\eta}{\lambda_{\min}(\tilde{\mathcal{A}}; D(\tilde{f}, J(\mathcal{U})d, \delta d))}\right\},
\end{aligned} \quad \text{(IV.7)}$$

and we want to lower-bound the quantity $\lambda_{\min}$:

$$\begin{aligned}
&\lambda_{\min}(\tilde{\mathcal{A}}; D(\tilde{f}, J(\mathcal{U})d, \delta d)) \\
&\qquad = \inf_{Y\in\tilde{E}(\mathcal{U})}\left[\sum_{i=1}^m |\mathrm{tr}(YJ(\mathcal{C}_i)d^2)|^2\right]^{1/2},
\end{aligned} \quad \text{(IV.8)}$$

where we define

$$\tilde{E}(\mathcal{U}) = \{Y \in D(\tilde{f}, J(\mathcal{U})d, \delta d) \mid \|Y\|_F = 1\}. \quad \text{(IV.9)}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

D. Mendelson's small-ball method

We will use Mendelson's small-ball method to lower-bound $\lambda_{\min}$. We will prove a uniform lower-bound that holds for all $\mathcal{U}$ simultaneously, i.e., we will lower-bound $\inf_{\mathcal{U}} \lambda_{\min}$. We define

$$\tilde{E} = \bigcup_{\mathcal{U}} \tilde{E}(\mathcal{U}), \tag{IV.10}$$

where the union runs over all unitary processes $\mathcal{U} : \rho \mapsto U\rho U^\dagger$ such that $U \in \mathbb{C}^{d\times d}$ satisfies the non-spikiness conditions (I.12). We then take the infimum over all $Y \in \tilde{E}$.

We will then lower-bound this quantity, using Mendelson's small-ball method (Theorem 8 in [9]): for any $\xi > 0$ and $t \geq 0$, with probability at least $1 - e^{-2t^2}$, we have that

$$\inf_{Y\in\tilde{E}} \left[\sum_{i=1}^m |\mathrm{tr}(YJ(\mathcal{C}_i)d^2)|^2\right]^{1/2} \geq \xi\sqrt{m}\tilde{Q}_{2\xi}(\tilde{E}) - 2\tilde{W}_m(\tilde{E}) - \xi t, \tag{IV.11}$$

where we define

$$\tilde{Q}_\xi(\tilde{E}) = \inf_{Y\in\tilde{E}} \left\{ \Pr_{\mathcal{C}\sim\tilde{G}} \left[|\mathrm{tr}(YJ(\mathcal{C})d^2)| \geq \xi\right]\right\}, \tag{IV.12}$$

$$\tilde{W}_m(\tilde{E}) = \mathop{\mathbb{E}}_{\substack{\epsilon_i\sim\pm1\\ \mathcal{C}\sim\tilde{G}}} \left[\sup_{Y\in\tilde{E}} \mathrm{tr}(Y\tilde{H})\right],$$

$$\text{with } \tilde{H} = \frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i J(\mathcal{C}_i)d^2, \tag{IV.13}$$

and $\varepsilon_1, \ldots, \varepsilon_m$ are Rademacher random variables.

E. Lower-bounding $\tilde{Q}_\xi(\tilde{E})$

We will lower-bound $\tilde{Q}_\xi(\tilde{E})$ as follows. Fix some $Y \in \tilde{E}$. We know that Y is in $\tilde{E}(\mathcal{U})$, for some unitary process $\mathcal{U} : \rho \mapsto U\rho U^\dagger$ such that U is non-spiky in the sense of (I.12). (As a result, $J(\mathcal{U})d$ also satisfies the last three constraints in (I.7).) Hence, we know that $\|Y\|_F = 1$, and there exists some $\tau > 0$ such that $J(\mathcal{U})d + \tau Y$ is non-spiky in the sense of (I.14) and satisfies the last three constraints in (I.7), and we have

$$\mathrm{tr}(J(\mathcal{U})d + \tau Y) \leq \mathrm{tr}(J(\mathcal{U})d), \quad \|\tau Y\|_F \geq \delta d. \tag{IV.14}$$

Note that we have $\|\tau Y\|_F = \tau \geq \delta d$.

We will lower-bound $\Pr[|\mathrm{tr}(YJ(\mathcal{C})d^2)| \geq \xi]$, for any $\xi \in [0, 1]$, using the Paley-Zygmund inequality, and appropriate bounds on the second and fourth moments of $\mathrm{tr}(YJ(\mathcal{C})d^2)$. To simplify the notation, let us define a random variable

$$\tilde{S} = \tau|\mathrm{tr}(YJ(\mathcal{C})d^2)|. \tag{IV.15}$$

1) Lower-bounding $\mathbb{E}(\tilde{S}^2)$: We will first put $\tilde{S}$ into a form that is easier to work with. We can write Y in the form $Y = J(\mathcal{Y})d$, which is the rescaled Choi matrix of some linear map $\mathcal{Y} \in \mathcal{L}(\mathbb{C}^{d\times d}, \mathbb{C}^{d\times d})$. Using the relationship between the trace of Choi and Liouville representations (see Appendix A), we have

$$\begin{aligned}\tilde{S} &= \tau|\,\mathrm{tr}(J(\mathcal{C})J(\mathcal{Y})d^3)|\\ &= \tau d|\,\mathrm{tr}((\mathcal{C}^L)^\dagger\mathcal{Y}^L)|\\ &= \tau d|\,\mathrm{tr}((C^\dagger\otimes C^T)\mathcal{Y}^L)|.\end{aligned} \tag{IV.16}$$

We can calculate $\mathbb{E}(\tilde{S}^2)$ by using the fact that $C \in \tilde{G}$ is chosen from a unitary 2-design:

$$\begin{aligned}\mathbb{E}(\tilde{S}^2) &= \tau^2 d^2 \int_{\text{Haar}} \left|\mathrm{tr}\left((U\otimes U^*\otimes U\otimes U^*)(\mathcal{Y}^L\otimes\mathcal{Y}^L)\right)\right| dU\\ &\geq \tau^2 d^2 \left|\mathrm{tr}\left(\int_{\text{Haar}} (U\otimes U^*\otimes U\otimes U^*)dU\,(\mathcal{Y}^L\otimes\mathcal{Y}^L)\right)\right|.\end{aligned} \tag{IV.17}$$

Now using Weingarten functions [31]–[33], we have that

$$\int_{\text{Haar}} (U\otimes U\otimes U^\dagger\otimes U^\dagger)dU = \frac{P_{3412}+P_{4321}}{d^2-1} - \frac{P_{3421}+P_{4312}}{d(d^2-1)} \tag{IV.18}$$

where

$$P_{\sigma(1),\sigma(2),\ldots\sigma(t)} = \sum_{j_1,j_2,\ldots,j_t} |j_1\rangle\langle j_{\sigma(1)}|\otimes|j_2\rangle\langle j_{\sigma(2)}|\otimes\cdots\otimes|j_t\rangle\langle j_{\sigma(t)}| \tag{IV.19}$$

is a permutation of the registers.

Let $\mathbb{T}_2(\cdot)$ transpose the second half of the indices of a matrix in $\mathbb{C}^{d^2\times d^2}$. That is, for any matrix

$$X = \sum_{ijkl} x_{ijkl}|i\rangle\langle j|\otimes|k\rangle\langle l|, \tag{IV.20}$$

we have

$$\mathbb{T}_2(X) = \sum_{ijkl} x_{ijkl}|i\rangle\langle j|\otimes|l\rangle\langle k|. \tag{IV.21}$$

Then note that

$$P_{1324}\mathbb{T}_2\left(\int_{\text{Haar}} (U\otimes U\otimes U^\dagger\otimes U^\dagger)dU\right)P_{1324} = \int_{\text{Haar}} (U\otimes U^*\otimes U\otimes U^*)dU. \tag{IV.22}$$

Combining with Eq. (IV.17) and Eq. (IV.18), we have

$$\mathbb{E}(\tilde{S}^2) \geq \tau^2 d^2 \left|\mathrm{tr}\left(P_{1324}\mathbb{T}_2\left(\frac{P_{3412}+P_{4321}}{d^2-1} - \frac{P_{3421}+P_{4312}}{d(d^2-1)}\right)P_{1324}(\mathcal{Y}^L\otimes\mathcal{Y}^L)\right)\right|. \tag{IV.23}$$

Because addition commutes with permutation and transposition and trace, we need to calculate

$$\mathrm{tr}\left(P_{1324}\mathbb{T}_2(P_\sigma)P_{1324}(\mathcal{Y}^L\otimes\mathcal{Y}^L)\right) \tag{IV.24}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

for $\sigma \in \{3412, 4312, 3421, 4321\}$. Letting $\sigma = (abcd)$, we have

$$\begin{aligned}
&\operatorname{tr}\left(P_{1324}\mathbb{T}_2(P_{abcd})P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) \\
&= \operatorname{tr}\left(P_{1324}\mathbb{T}_2\left(\sum_{j_1,j_2,j_3,j_4} |j_1\rangle\langle j_a| \otimes |j_2\rangle\langle j_b| \otimes |j_3\rangle\langle j_c| \otimes |j_4\rangle\langle j_d|\right) P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) \\
&= \operatorname{tr}\left(P_{1324}\sum_{j_1,j_2,j_3,j_4} |j_1\rangle\langle j_a| \otimes |j_2\rangle\langle j_b| \otimes |j_c\rangle\langle j_3| \otimes |j_d\rangle\langle j_4| P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) \\
&= \operatorname{tr}\left(\sum_{j_1,j_2,j_3,j_4} |j_1\rangle\langle j_a| \otimes |j_c\rangle\langle j_3| \otimes |j_2\rangle\langle j_b| \otimes |j_d\rangle\langle j_4| (\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) \\
&= \sum_{j_1,j_2,j_3,j_4} \langle j_a j_3|\mathcal{Y}^L|j_1 j_c\rangle\langle j_b j_4|\mathcal{Y}^L|j_2 j_d\rangle. \qquad \text{(IV.25)}
\end{aligned}$$

We know that $Y = J(\mathcal{Y})d$ satisfies the last three constraints in (I.7). Because of these constraints, we have

$$\begin{aligned}
\sum_{i_1,i_2,i_3} \langle i_2 i_3|\mathcal{Y}^L|i_1 i_1\rangle &= \sum_{i_1,i_2,i_3} d\langle i_2 i_1|J(\mathcal{Y})|i_3 i_1\rangle \\
&= \sum_{i_2,i_3} d\langle i_2|\operatorname{tr}_2(J(\mathcal{Y}))|i_3\rangle \qquad \text{(IV.26)} \\
&= d\operatorname{tr}(J(\mathcal{Y})),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i_1,i_2,i_3} \langle i_1 i_1|\mathcal{Y}^L|i_2 i_3\rangle &= \sum_{i_1,i_2,i_3} d\langle i_1 i_2|J(\mathcal{Y})|i_1 i_3\rangle \\
&= \sum_{i_2,i_3} d\langle i_2|\operatorname{tr}_1(J(\mathcal{Y}))|i_3\rangle \qquad \text{(IV.27)} \\
&= d\operatorname{tr}(J(\mathcal{Y})).
\end{aligned}$$

We now go through the four permutations of Eq. (IV.23):

- $\mathbf{P_{3412}}$. In this case, $a = 3$, $b = 4$, $c = 1$, and $d = 2$. Plugging into Eq. (IV.25) and using Eqs. (IV.27) and (IV.26), we have

$$\operatorname{tr}\left(P_{1324}\mathbb{T}_2(P_{3412})P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) = d^2\operatorname{tr}(J(\mathcal{Y}))^2. \qquad \text{(IV.28)}$$

- $\mathbf{P_{4321}}$. In this case, $a = 4$, $b = 3$, $c = 2$, and $d = 1$. Plugging into Eq. (IV.25), we have

$$\begin{aligned}
&\operatorname{tr}\left(P_{1324}\mathbb{T}_2(P_{4321})P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) \\
&= \sum_{j_1,j_2,j_3,j_4} \langle j_4 j_3|\mathcal{Y}^L|j_1 j_2\rangle\langle j_3 j_4|\mathcal{Y}^L|j_2 j_1\rangle \\
&= \sum_{j_1,j_2,j_3,j_4} \langle j_4 j_3|\mathcal{Y}^L|j_1 j_2\rangle\langle j_4 j_3|(\mathcal{Y}^L)^*|j_1 j_2\rangle \\
&= \sum_{j_1,j_2,j_3,j_4} |\langle j_4 j_3|\mathcal{Y}^L|j_1 j_2\rangle|^2 \\
&= \|\mathcal{Y}^L\|_F^2 \\
&= d^2\|J(\mathcal{Y})\|_F^2, \qquad \text{(IV.29)}
\end{aligned}$$

where we used Eq. (A.4).

- $\mathbf{P_{3421}}$. In this case, $a = 3$, $b = 4$, $c = 2$, and $d = 1$. Plugging into Eq. (IV.25) and using Eqs. (IV.27) and (IV.26), we have

$$\operatorname{tr}\left(P_{1324}\mathbb{T}_2(P_{3412})P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) = d^2\operatorname{tr}(J(\mathcal{Y}))^2. \qquad \text{(IV.30)}$$

- $\mathbf{P_{4312}}$. In this case, $a = 4$, $b = 3$, $c = 1$, and $d = 2$. Plugging into Eq. (IV.25) and using Eqs. (IV.27) and (IV.26), we have

$$\operatorname{tr}\left(P_{1324}\mathbb{T}_2(P_{3412})P_{1324}(\mathcal{Y}^L \otimes \mathcal{Y}^L)\right) = d^2\operatorname{tr}(J(\mathcal{Y}))^2. \qquad \text{(IV.31)}$$

Putting it all together, we have

$$\begin{aligned}
\mathbb{E}(\tilde{S}^2) &\geq \tau^2 d^4 \left|\frac{\operatorname{tr}(J(\mathcal{Y}))^2 + \|J(\mathcal{Y})\|_F^2}{d^2-1} - \frac{2\operatorname{tr}(J(\mathcal{Y}))^2}{d(d^2-1)}\right| \\
&\geq \tau^2 d^2(1-\tfrac{2}{d})\operatorname{tr}(J(\mathcal{Y}))^2 + \tau^2 d^2\|J(\mathcal{Y})\|_F^2 \\
&\geq 0 + \tau^2 d^2\|J(\mathcal{Y})\|_F^2 = \tau^2\|Y\|_F^2 = \tau^2. \qquad \text{(IV.32)}
\end{aligned}$$

2) Upper-bounding $\mathbb{E}(\tilde{S}^4)$: To handle $\mathbb{E}(\tilde{S}^4)$, we need to use a different argument from that in [9], since we are sampling from a unitary 2-design, not a 4-design. Our solution is to upper-bound $\tilde{S}$, using the non-spikiness properties of $J(\mathcal{U})d$ and $J(\mathcal{U})d + \tau Y$:

$$\begin{aligned}
\tilde{S} &= \left|-\operatorname{tr}(J(\mathcal{U})dJ(\mathcal{C})d^2) + \operatorname{tr}((J(\mathcal{U})d + \tau Y)J(\mathcal{C})d^2)\right| \\
&\leq \beta d. \qquad \text{(IV.33)}
\end{aligned}$$

This implies an upper-bound on $\mathbb{E}(\tilde{S}^4)$:

$$\mathbb{E}(\tilde{S}^4) \leq \beta^2 d^2 \mathbb{E}(\tilde{S}^2). \qquad \text{(IV.34)}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

Putting it all together, and using the Paley-Zygmund inequality, for any $\xi \in [0,1]$, we have:

$$\begin{aligned}\Pr[|\operatorname{tr}(YJ(\mathcal{C})d^2)| \geq \xi] &= \Pr[\tilde{S}^2 \geq \xi^2\tau^2] \\ &\geq \Pr[\tilde{S}^2 \geq \xi^2\mathbb{E}(\tilde{S}^2)] \\ &\geq (1-\xi^2)^2\frac{\mathbb{E}(\tilde{S}^2)^2}{\mathbb{E}(\tilde{S}^4)} \\ &\geq (1-\xi^2)^2\frac{\tau^2}{\beta^2 d^2}.\end{aligned} \tag{IV.35}$$

Thus, using the fact that $\tau \geq \delta$, we have that

$$\tilde{Q}_\xi(\tilde{E}) \geq (1-\xi^2)^2\frac{\delta^2}{\beta^2}. \tag{IV.36}$$

F. Upper-bounding $\tilde{W}_m(E)$

In this section, we will bound

$$\tilde{W}_m(E) = \mathbb{E}\sup_{Y\in\tilde{E}}\operatorname{tr}(Y\tilde{H}), \quad \tilde{H} = \frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i J(\mathcal{C}_i)d^2, \tag{IV.37}$$

where the ε_i are Rademacher random variables, and the expectation is taken both over the ε_i and the choice of the unitaries C_i. For this bound, we will only need the fact that the C_i are chosen from a unitary 1-design, rather than a unitary 2-design.

We start by following the argument in [9], where it is shown that

$$\mathbb{E}\sup_{Y\in F_r}\operatorname{tr}(Y\tilde{H}) \leq \mathbb{E}2\sqrt{r}\|\tilde{H}\| \tag{IV.38}$$

where

$$F_r = \bigcup_{X_{\text{true}}} F(X_{\text{true}}), \tag{IV.39}$$

and the union is over all $X_{\text{true}} \in \mathbb{C}^{d\times d}_{\text{Herm}}$ with rank at most r, and

$$F(X_{\text{true}}) = \{Y \in D(\|\cdot\|_*, X_{\text{true}}, 0) \mid \|Y\|_F = 1\}. \tag{IV.40}$$

This can be compared with our definition of the set $\tilde{E}$ in (IV.10) and (IV.9).

In our case, we have $\tilde{E} = \bigcup_{\mathcal{U}} \tilde{E}(\mathcal{U})$, where the union is over all processes $\mathcal{U}$ whose unnormalized Choi state $J(\mathcal{U})d \in \mathbb{C}^{d^2\times d^2}_{\text{Herm}}$ has rank 1, and where $\mathcal{U}$ satisfies the non-spikiness condition (I.14). $\tilde{E}(\mathcal{U})$ is defined similarly to $F(X_{\text{true}})$, but using the function f from (IV.6) rather than the trace norm. While f involves the trace $\operatorname{tr}(\cdot)$ instead of the trace norm $\|\cdot\|_*$, because we are considering only positive semidefinite matrices, it can be replaced by the trace norm. Hence $\tilde{E} \subset F_1$, and so

$$\mathbb{E}\sup_{Y\in\tilde{E}}\operatorname{tr}(Y\tilde{H}) \leq \mathbb{E}\sup_{Y\in F_1}\operatorname{tr}(Y\tilde{H}) \leq 2\mathbb{E}\|\tilde{H}\|. \tag{IV.41}$$

Now we analyze

$$\mathbb{E}_{\varepsilon,\mathcal{C}}\left\|\frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i J(\mathcal{C}_i)d^2\right\|, \tag{IV.42}$$

using similar tools to what is used in [9]. Since the ϵ_j's form a Rademacher sequence and $J(\mathcal{C}_i)$ are Hermitian, we can apply the non-commutative Khintchine inequality [55], [57]:

$$\begin{aligned}&\mathbb{E}_{\varepsilon,\mathcal{C}}\left\|\frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i J(\mathcal{C}_i)d^2\right\| \\ &\quad\leq \mathbb{E}_{\mathcal{C}}\sqrt{2\ln(2d^2)/m}\left\|\sum_{i=1}^m J(\mathcal{C}_i)^2\right\|^{1/2} d^2 \\ &\quad\leq \sqrt{2\ln(2d^2)/m}\left(\mathbb{E}_{\mathcal{C}}\left\|\sum_{i=1}^m J(\mathcal{C}_i)\right\|\right)^{1/2} d^2,\end{aligned} \tag{IV.43}$$

where in the second line we used Jensen's inequality, and we used the fact that $J(\mathcal{C}_i)$ is a rank one projector with trace 1, so $J(\mathcal{C}_i)^2 = J(\mathcal{C}_i)$.

We now apply the matrix Chernoff inequality of Theorem 15 in [9] to $\mathbb{E}_{\mathcal{C}}\|\sum_{i=1}^m J(\mathcal{C}_i)\|$. To apply the theorem, we need to bound

$$\left\|\mathbb{E}_{\mathcal{C}}\sum_{i=1}^m J(\mathcal{C}_i)\right\| \quad \text{and} \quad \|J(\mathcal{C}_i)\|. \tag{IV.44}$$

Since $J(\mathcal{C}_i)$ corresponds to a quantum state, its maximum eigenvalue is 1, so $\|J(\mathcal{C}_i)\| = 1$. Next, notice

$$\mathbb{E}_{\mathcal{C}}\sum_{i=1}^m J(\mathcal{C}_i) = \sum_{i=1}^m \mathbb{E}_{\mathcal{C}} J(\mathcal{C}_i) = \frac{m}{d^2}I. \tag{IV.45}$$

Because the C_i are drawn from a unitary 2-design, $\mathbb{E}_{\mathcal{C}}J(\mathcal{C}_i)$ is the state that results when a depolarizing channel is applied to one half of a maximally entangled state. (Randomly applying operations from a unitary 1-design results in the depolarizing channel.) The resulting state is the maximally mixed state I/d^2 in $\mathbb{C}^{d^2\times d^2}$. Thus

$$\left\|\mathbb{E}_{\mathcal{C}}\sum_{i=1}^m J(\mathcal{C}_i)\right\| = \frac{m}{d^2}. \tag{IV.46}$$

Then the Matrix Chernoff Inequality (Theorem 15 in [9]) tells us that for any $\gamma > 0$,

$$\mathbb{E}\left\|\sum_{j=1}^m J(\mathcal{C}_i)\right\| \leq \frac{(e^\gamma - 1)m + d^2\log d^2}{d^2\gamma}. \tag{IV.47}$$

Minimizing γ gives

$$\mathbb{E}\left\|\sum_{j=1}^m J(\mathcal{C}_i)\right\| \leq c_4 m/d^2 \tag{IV.48}$$

for some numerical constant $c_4 > 0$. Plugging this expression into Eq. (IV.43), we obtain the desired bound:

$$\tilde{W}_m(\tilde{E}) \leq c_5\sqrt{\ln d}\cdot d, \tag{IV.49}$$

where $c_5 > 0$ is some numerical constant.

G. Final result

Combining equations (IV.8), (IV.11), (IV.36) and (IV.49), and setting $\xi = 1/4$, $\nu = \beta/\delta$ and $t = \frac{1}{16}\sqrt{m}\nu^{-2}$, we get that:

$$\begin{aligned}&\inf_{\mathcal{U}} \lambda_{\min}(\tilde{\mathcal{A}}; D(\tilde{f}, J(\mathcal{U})d, \delta d)) \\ &\quad \geq \frac{1}{4}\sqrt{m}\,\frac{9}{16\nu^2} - 2c_5\sqrt{\ln d}\cdot d - \frac{1}{4}\,t \\ &\quad = \frac{1}{8}\sqrt{m}\,\frac{1}{\nu^2} - 2c_5\sqrt{\ln d}\cdot d.\end{aligned} \tag{IV.50}$$

Now we set $m \geq \left(8\nu^2 c_0\right)^2 \cdot d^2 \ln d$, which implies

$$\frac{1}{8\nu^2 c_0}\sqrt{m} \geq \sqrt{\ln d}\cdot d, \tag{IV.51}$$

hence

$$\begin{aligned}&\inf_{\mathcal{U}} \lambda_{\min}(\tilde{\mathcal{A}}; D(\tilde{f}, J(\mathcal{U})d, \delta d)) \\ &\quad \geq \frac{\sqrt{m}}{8\nu^2}\left(1 - \frac{2c_5}{c_0}\right).\end{aligned} \tag{IV.52}$$

This can be plugged into our approximate recovery bound (IV.7). This finishes the proof of Theorem I.3. □

V. Phase Retrieval Using Unitary 4-Designs

In this section, we again consider the PhaseLift algorithm for recovering an unknown unitary matrix $U \in \mathbb{C}^{d\times d}$. However, we now consider a stronger measurement ensemble: we let the measurement matrices $C_1, \ldots, C_m$ be chosen from a unitary 4-design $\hat{G}$ in $\mathbb{C}^{d\times d}$. This leads to a stronger recovery guarantee: PhaseLift achieves *exact* recovery of *all* unitaries U. This claim was presented in Theorem I.1. We show the proof here.

We use the same notation as in the previous section, except that here we call the measurement operator $\hat{\mathcal{A}}$ rather than $\tilde{\mathcal{A}}$. We can write the desired solution of the PhaseLift convex program as $J(\mathcal{U})d$, and we can write the measurement operator $\hat{\mathcal{A}}$ as

$$\hat{\mathcal{A}}(\Gamma) = \left[\operatorname{tr}(\Gamma J(\mathcal{C}_i)d^2)\right]_{i=1}^{m}. \tag{V.1}$$

We want to recover the solution $J(\mathcal{U})d$, up to an additive error of size $\delta\|J(\mathcal{U})d\|_F = \delta d$.

A. Descent cone

For the case of unitary 4-designs, we use a similar descent cone to the one used in [9], [30] — that is, for $f : \mathbb{C}^{d^2\times d^2}_{\text{Herm}} \to \mathbb{R}\cup\{\infty\}$ a proper convex function, and $X_{\text{true}} \in \mathbb{C}^{d^2\times d^2}_{\text{Herm}}$, we use the descent cone $D(f, X_{\text{true}}, 0)$. This is the set of all directions, originating at the point X_{true}, that cause the value of f to decrease.

We define the function $\hat{f} : \mathbb{C}^{d^2\times d^2}_{\text{Herm}} \to \mathbb{R}\cup\{\infty\}$ as follows:

$$\hat{f}(\Gamma) = \begin{cases} \operatorname{tr}(\Gamma) & \text{if } \Gamma \text{ satisfies the} \\ & \text{last three constraints in (I.7),} \\ \infty & \text{otherwise.}\end{cases} \tag{V.2}$$

By Prop. 7 in [9], our recovery bound is:

$$\|\Gamma_{\text{opt}} - J(\mathcal{U})d\|_F \leq \frac{2\eta}{\lambda_{\min}(\hat{\mathcal{A}}; D(\hat{f}, J(\mathcal{U})d, 0))}, \tag{V.3}$$

and we want to lower-bound the quantity $\lambda_{\min}$:

$$\lambda_{\min}(\hat{\mathcal{A}}; D(\hat{f}, J(\mathcal{U})d, 0)) = \inf_{Y\in\hat{E}(\mathcal{U})}\left[\sum_{i=1}^{m}|\operatorname{tr}(YJ(\mathcal{C}_i)d^2)|^2\right]^{1/2}, \tag{V.4}$$

where we define

$$\hat{E}(\mathcal{U}) = \{Y \in D(\hat{f}, J(\mathcal{U})d, 0) \mid \|Y\|_F = 1\}. \tag{V.5}$$

B. Mendelson's small-ball method

We will use Mendelson's small-ball method to lower-bound $\lambda_{\min}$. We will prove a uniform lower-bound that holds for all $\mathcal{U}$ simultaneously, i.e., we will lower-bound $\inf_{\mathcal{U}} \lambda_{\min}$. We define

$$\hat{E} = \bigcup_{\mathcal{U}} \hat{E}(\mathcal{U}), \tag{V.6}$$

where the union runs over all unitary processes $\mathcal{U} \in \mathcal{L}(\mathbb{C}^{d\times d}, \mathbb{C}^{d\times d})$ of the form $\mathcal{U} : \rho \mapsto U\rho U^\dagger$. We then take the infimum over all $Y \in \hat{E}$.

We will then lower-bound this quantity, using Mendelson's small-ball method (Theorem 8 in [9]): for any $\xi > 0$ and $t \geq 0$, with probability at least $1 - e^{-2t^2}$, we have that

$$\begin{aligned}&\inf_{Y\in\hat{E}}\left[\sum_{i=1}^{m}|\operatorname{tr}(YJ(\mathcal{C}_i)d^2)|^2\right]^{1/2} \\ &\qquad \geq \xi\sqrt{m}\hat{Q}_{2\xi}(\hat{E}) - 2\hat{W}_m(\hat{E}) - \xi t,\end{aligned} \tag{V.7}$$

where we define

$$\hat{Q}_\xi(\hat{E}) = \inf_{Y\in\hat{E}}\left\{\Pr_{C\sim\hat{G}}\left[|\operatorname{tr}(YJ(\mathcal{C})d^2)| \geq \xi\right]\right\}, \tag{V.8}$$

$$\hat{W}_m(\hat{E}) = \mathop{\mathbb{E}}_{\substack{\epsilon_i\sim\pm1\\ C_i\sim\hat{G}}}\left[\sup_{Y\in\hat{E}}\operatorname{tr}(Y\hat{H})\right], \quad \hat{H} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varepsilon_i J(\mathcal{C}_i)d^2, \tag{V.9}$$

and $\varepsilon_1, \ldots, \varepsilon_m$ are Rademacher random variables.

C. Lower-bounding $\hat{Q}_\xi(\hat{E})$

We will lower-bound $\Pr_{C\sim\hat{G}}[|\operatorname{tr}(YJ(\mathcal{C})d^2)| \geq \xi]$, for any $\xi \in [0,1]$, using the Paley-Zygmund inequality, and appropriate bounds on the second and fourth moments of $|\operatorname{tr}(YJ(\mathcal{C})d^2)|$. To simplify the notation, let us define a random variable

$$\hat{S} = |\operatorname{tr}(YJ(\mathcal{C})d^2)|. \tag{V.10}$$

We will first put $\hat{S}$ into a form that is easier to work with. We can write Y in the form $Y = J(\mathcal{Y})d$, which is the rescaled Choi matrix of some linear map $\mathcal{Y} \in \mathcal{L}(\mathbb{C}^{d\times d}, \mathbb{C}^{d\times d})$. Using the relationship between the trace of Choi and Liouville representations (see Appendix A), we have

$$\begin{aligned}\hat{S} &= |\operatorname{tr}(J(\mathcal{Y})dJ(\mathcal{C})d^2)| \\ &= d|\operatorname{tr}((\mathcal{C}^L)^\dagger\mathcal{Y}^L)| \\ &= d|\operatorname{tr}((C^\dagger\otimes C^T)\mathcal{Y}^L)|.\end{aligned} \tag{V.11}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

1) Lower-bounding $\mathbb{E}(\hat{S}^2)$: Because we are working with a unitary 4-design, and $\hat{S}^2$ only depends on the second moment of the distribution, the bound will be the same as for a unitary 2-design, and we can use our bound from Section IV-E1, Eq. (IV.32) (with $\tau = 1$):

$$\begin{aligned}\mathbb{E}(\hat{S}^2) &\geq d^2(1-\tfrac{2}{d})\operatorname{tr}(J(\mathcal{Y}))^2 + d^2\|J(\mathcal{Y})\|_F^2 \\ &\geq \tfrac{1}{2}d^2 \max\left\{\operatorname{tr}(J(\mathcal{Y}))^2, \|J(\mathcal{Y})\|_F^2\right\}, \end{aligned} \tag{V.12}$$

and also (using Eq. (IV.32)):

$$\mathbb{E}(\hat{S}^2) \geq d^2\|J(\mathcal{Y})\|_F^2 = \|Y\|_F^2 = 1. \tag{V.13}$$

2) Upper-bounding $\mathbb{E}(\hat{S}^4)$: Now we would like to bound $\mathbb{E}[\hat{S}^4]$. Because we are considering a unitary 4-design, which has the same fourth moments as the Haar measure on unitaries, instead of taking the average $\mathbb{E}[\hat{S}^4]$ over the 4-design, we will take the average over Haar random unitaries.

We have

$$\begin{aligned}\mathbb{E}[\hat{S}^4] =& d^4 \int_{Haar} dU |\operatorname{tr}\left((U\otimes U^*)\mathcal{Y}^L\right)|^4 \\ =& d^4 \int_{Haar} dU \\ & \left(\operatorname{tr}\left((U\otimes U^*)\mathcal{Y}^L\right)\operatorname{tr}\left((U^*\otimes U)(\mathcal{Y}^L)^*\right)\right)^2 \\ =& d^4 \int_{Haar} dU \\ & \operatorname{tr}\left(\left(U\otimes (U^*)^{\otimes 2}\otimes U\right)^{\otimes 2}\left(\mathcal{Y}^L\otimes(\mathcal{Y}^L)^*\right)^{\otimes 2}\right). \end{aligned} \tag{V.14}$$

Using similar tricks as in the case of the second moment, we have that

$$\begin{aligned}&\int_{Haar} dU \left(U\otimes (U^*)^{\otimes 2}\otimes U\right)^{\otimes 2} \\ &= P_{15842673}\mathbb{T}_2\left(\int_{Haar} dU U^{\otimes 4}\otimes (U^\dagger)^{\otimes 4}\right) P_{15842673}. \end{aligned} \tag{V.15}$$

We will define $P^* = P_{15842673}$. Then we use Weingarten functions [31]–[33] to obtain an expression for the integral on the right hand side of Eq. (V.15) in terms of permutation operators:

$$\int_{Haar} dU U^{\otimes 4}\otimes (U^\dagger)^{\otimes 4} = \sum_{\sigma,\tau\in\mathbb{S}_4} W_g(d,4,\sigma\tau^{-1}) P_{\sigma,\tau}, \tag{V.16}$$

where $\mathbb{S}_4$ is the symmetric group of 4 elements, and

$$\begin{aligned}&P_{\sigma,\tau} = \\ &P_{\tau(1)+4,\tau(2)+4,\tau(3)+4,\tau(4)+4,\sigma^{-1}(1),\sigma^{-1}(2),\sigma^{-1}(3),\sigma^{-1}(4)}. \end{aligned} \tag{V.17}$$

Combining Eqs, (V.14), (V.15), and (V.16), we have

$$\begin{aligned}\mathbb{E}[\hat{S}^4] = d^4 &\sum_{\sigma,\tau\in\mathbb{S}_4} W_g(d,4,\sigma\tau^{-1})\times \\ &\operatorname{tr}\left(P^*\mathbb{T}_2(P_{\sigma,\tau})P^*\left(\mathcal{Y}^L\otimes(\mathcal{Y}^L)^*\right)^{\otimes 2}\right). \end{aligned} \tag{V.18}$$

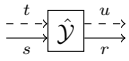
The permutations $P_{\sigma,\tau}$ in the sum will be of the form $P_{wxyzabcd}$ (where $(wxyz)$ is a nonrepeating sequence of elements in the set $\{5,6,7,8\}$ and $(abcd)$ is a non-repeating sequence of elements in the set $\{1,2,3,4\}$. Each term in the above sum is therefore of the form:

$$
\begin{aligned}
&\operatorname{tr}\left(P^*\mathbb{T}_2\left(P_{wxyzabcd}\right)P^*\left(\mathcal{Y}^L\otimes(\mathcal{Y}^L)^*\right)^{\otimes 2}\right)\\
&=\operatorname{tr}\left(P^*\mathbb{T}_2\left(\sum_{\substack{j_1,j_2,j_3,j_4\\ j_5,j_6,j_7,j_8}}|j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\rangle\langle j_w,j_x,j_y,j_z,j_a,j_b,j_c,j_d|\right)P^*\left(\mathcal{Y}^L\otimes(\mathcal{Y}^L)^*\right)^{\otimes 2}\right)\\
&=\operatorname{tr}\left(P^*\sum_{\substack{j_1,j_2,j_3,j_4\\ j_5,j_6,j_7,j_8}}|j_1,j_2,j_3,j_4,j_a,j_b,j_c,j_d\rangle\langle j_w,j_x,j_y,j_z,j_5,j_6,j_7,j_8|P^*\left(\mathcal{Y}^L\otimes(\mathcal{Y}^L)^*\right)^{\otimes 2}\right)\\
&=\operatorname{tr}\left(\sum_{\substack{j_1,j_2,j_3,j_4\\ j_5,j_6,j_7,j_8}}|j_1,j_a,j_d,j_4,j_2,j_b,j_c,j_3\rangle\langle j_w,j_5,j_8,j_z,j_x,j_6,j_7,j_y|\left(\mathcal{Y}^L\otimes(\mathcal{Y}^L)^*\right)^{\otimes 2}\right)\\
&=\sum_{\substack{j_1,j_2,j_3,j_4\\ j_5,j_6,j_7,j_8}}\langle j_w,j_5|\mathcal{Y}^L|j_1,j_a\rangle\langle j_8,j_z|(\mathcal{Y}^L)^*|j_d,j_4\rangle\langle j_x,j_6|\mathcal{Y}^L|j_2,j_b\rangle\langle j_7,j_y|(\mathcal{Y}^L)^*|j_c,j_3\rangle\\
&=\sum_{\substack{j_1,j_2,j_3,j_4\\ j_5,j_6,j_7,j_8}}\langle j_w,j_5|\mathcal{Y}^L|j_1,j_a\rangle\langle j_z,j_8|\mathcal{Y}^L|j_4,j_d\rangle\langle j_x,j_6|\mathcal{Y}^L|j_2,j_b\rangle\langle j_y,j_7|\mathcal{Y}^L|j_3,j_c\rangle\\
&=\sum_{\substack{j_1,j_2,j_3,j_4\\ j_5,j_6,j_7,j_8}}\langle j_1,j_5|\hat{\mathcal{Y}}|j_a,j_w\rangle\langle j_4,j_8|\hat{\mathcal{Y}}|j_d,j_z\rangle\langle j_2,j_6|\hat{\mathcal{Y}}|j_b,j_x\rangle\langle j_3,j_7|\hat{\mathcal{Y}}|j_c,j_y\rangle
\end{aligned}
\tag{V.19}
$$

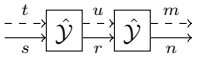
where in the second to last line we have used Eq. (A.4), and in the last line, we have reordered the elements of $\mathcal{Y}^L$ as

$$\langle j_r,j_s|\mathcal{Y}^L|j_t,j_u\rangle=\langle j_t,j_s|\hat{\mathcal{Y}}|j_u,j_r\rangle. \tag{V.20}$$

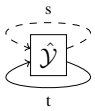
Now we will use a graphical representation of the matrices $\hat{\mathcal{Y}}$ in order to evaluate these terms:



$$= \quad \langle j_t,j_s|\hat{\mathcal{Y}}|j_u,j_r\rangle \tag{V.21}$$

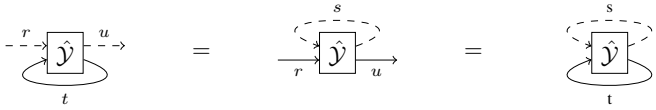


$$= \quad \langle j_t,j_s|\hat{\mathcal{Y}}\hat{\mathcal{Y}}|j_m,j_n\rangle \quad = \sum_{u,r}\langle j_t,j_s|\hat{\mathcal{Y}}|j_u,j_r\rangle\langle j_u,j_r|\hat{\mathcal{Y}}|j_m,j_n\rangle \tag{V.22}$$



$$= \quad \sum_{j_s,j_t}\langle j_t,j_s|\hat{\mathcal{Y}}|j_t,j_s\rangle=\operatorname{tr}(\hat{\mathcal{Y}}) \tag{V.23}$$

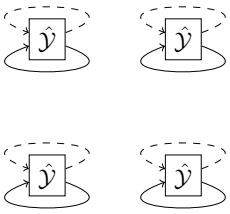
Furthermore, because of Eq. (IV.26) and Eq. (IV.27),



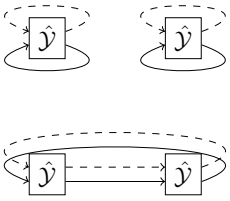
$$= \quad d\operatorname{tr}(J(\mathcal{Y})). \tag{V.24}$$

We see that a single self-loop on either register forces the other register to also have a self loop. Because of this simplifying effect, we can enumerate all possible configurations of commutative diagrams that arise from choices of $(wxyzabcd)$ in the last line of Eq. (V.19). We have the following 7 diagrams that can represent the last line of Eq. (V.19):

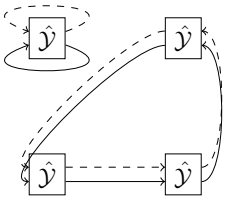
I :



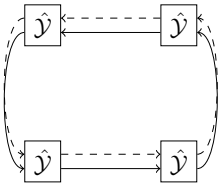
II :



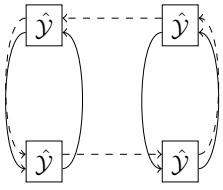
III :



IV :

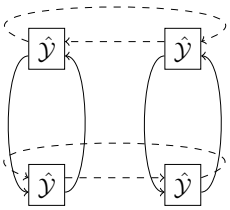


V :



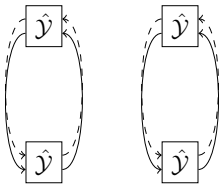
(V.25)

VI :



(V.26)

VII :



(V.27)

These commutative diagrams were found in the following way. First, there is the case that all four $\hat{\mathcal{Y}}$'s have self loops. This corresponds to Diagram I. Next, we consider the case that three $\hat{\mathcal{Y}}$'s have self loops, but in this case, the only way to connect up the final $\hat{\mathcal{Y}}$'s tensor legs is to create self loops, so we are back to Diagram I. In the case that two $\hat{\mathcal{Y}}$'s have self loops, the only way to connect the remaining two $\hat{\mathcal{Y}}$'s without giving them self loops is as shown in Diagram II. Continuing in this way, we can enumerate all possible diagrams.

We ignore diagrams that are identical to diagrams that are depicted, but which have one or more loops reversed in direction, or that have dashed and solid arrows switched. This is acceptable, because by reordering the elements of the matrix, as we did in Eq. (V.20), we can create a new figure which looks identical to ones shown above, but involving a new matrix $\hat{\mathcal{Y}}'$ whose elements are the same as $\hat{\mathcal{Y}}$. In our analysis below, we will ultimately see that our bounds on the contributions due to each figure will depend only on the Frobenius norm of $\mathcal{Y}^L$, which only depends on the elements of the matrix, and not on their ordering. (We also have contributions that depend on the $\mathrm{tr}(\mathcal{Y})$, but these terms only come from self-loops about a single element, for which reordering does not produce a new term.)

For a square matrix A and integer $i > 1$, we will use the following bound:

$$\begin{aligned} \mathrm{tr}(A^i) &= \mathrm{tr}(A^{i-1}A) \\ &= \mathrm{vec}((A^{i-1})^T)^T \mathrm{vec}(A) \\ &\leq \|(A^{i-1})^T\|_F \|A\|_F \\ &\leq \|A\|_F^i \end{aligned} \tag{V.28}$$

where we have used Cauchy-Schwarz, the submultiplicative property of the Frobenius norm, and the fact that $\|A^T\|_F = \|A\|_F$.

We now bound the contribution due to each diagram.

I: We read off the contribution of $\mathrm{tr}(\hat{\mathcal{Y}})^4 = d^4 \,\mathrm{tr}(J(\mathcal{Y}))^4$.

II: We have a contribution of

$$\begin{aligned} d^2 \operatorname{tr}(J(\mathcal{Y}))^2 \operatorname{tr}(\hat{\mathcal{Y}}^2) \leq & d^2 \operatorname{tr}(J(\mathcal{Y}))^2 \|\hat{\mathcal{Y}}\|_F^2 \\ = & d^4 \operatorname{tr}(J(\mathcal{Y}))^2 \|J(\mathcal{Y})\|_F^2. \end{aligned} \tag{V.29}$$

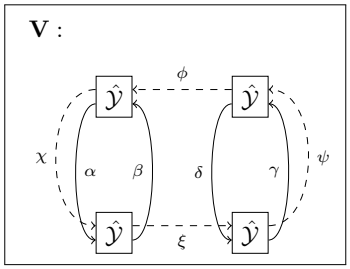
III: We have a contribution of

$$\begin{aligned} d \operatorname{tr}(J(\mathcal{Y})) \operatorname{tr}(\hat{\mathcal{Y}}^3) \leq & d \operatorname{tr}(J(\mathcal{Y})) \|\hat{\mathcal{Y}}^2\|_F^3 \\ = & d^4 \operatorname{tr}(J(\mathcal{Y})) \|J(\mathcal{Y})\|_F^3. \end{aligned} \tag{V.30}$$

IV: We have a contribution of

$$\operatorname{tr}(\hat{\mathcal{F}}^4) \leq d^4 \|J(\mathcal{F})\|_F^4. \tag{V.31}$$

V: We reprint the diagram here, but with labels:



(V.32)

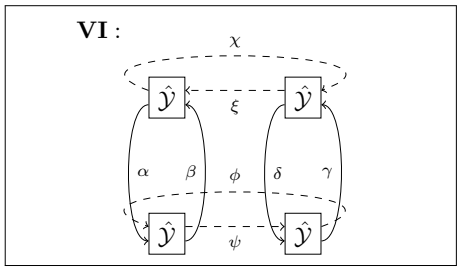
Let K be the $d \times d$ matrix such that

$$\langle j_a | K | j_b \rangle = \sqrt{\sum_{j_c, j_d} \left(\langle j_a, j_c | \hat{\mathcal{Y}} | j_b, j_d \rangle \right)^2}. \tag{V.33}$$

Then the contribution of the diagram is bounded by

$$\begin{aligned} & \sum_{\substack{j_\alpha, j_\beta, j_\gamma, j_\delta \\ j_\xi, j_\chi, j_\phi, j_\psi}} \langle j_\phi, j_\beta | \hat{\mathcal{Y}} | j_\chi, j_\alpha \rangle \langle j_\chi, j_\alpha | \hat{\mathcal{Y}} | j_\xi, j_\beta \rangle \\ & \qquad \times \langle j_\xi, j_\delta | \hat{\mathcal{Y}} | j_\psi, j_\gamma \rangle \langle j_\psi, j_\gamma | \hat{\mathcal{Y}} | j_\phi, j_\delta \rangle \\ & \leq \sum_{\substack{j_\xi, j_\chi \\ j_\phi, j_\psi}} \sqrt{\sum_{j_\alpha, j_\beta} \left(\langle j_\phi, j_\beta | \hat{\mathcal{Y}} | j_\chi, j_\alpha \rangle \right)^2} \\ & \qquad \times \sqrt{\sum_{j_\alpha, j_\beta} \left(\langle j_\chi, j_\alpha | \hat{\mathcal{Y}} | j_\xi, j_\beta \rangle \right)^2} \\ & \qquad \times \sqrt{\sum_{j_\gamma, j_\delta} \left(\langle j_\xi, j_\delta | \hat{\mathcal{Y}} | j_\psi, j_\gamma \rangle \right)^2} \\ & \qquad \times \sqrt{\sum_{j_\gamma, j_\delta} \left(\langle j_\psi, j_\gamma | \hat{\mathcal{Y}} | j_\phi, j_\delta \rangle \right)^2} \\ & \leq \sum_{\substack{j_\xi, j_\chi \\ j_\phi, j_\psi}} \langle j_\phi | K | j_\chi \rangle \langle j_\chi | K | j_\xi \rangle \langle j_\xi | K | j_\psi \rangle \langle j_\psi | K | j_\phi \rangle \\ & \qquad = \operatorname{tr}(K^4) \\ & \qquad \leq \|K\|_F^4 \\ & \qquad = \left(\sum_{j_a, j_b} \sum_{j_c, j_d} \left(\langle j_a, j_c | \hat{\mathcal{Y}} | j_b, j_d \rangle \right)^2 \right)^2 \\ & \qquad = \|\hat{\mathcal{Y}}\|_F^4 \\ & \qquad = d^4 \|J(\mathcal{Y})\|_F^4. \end{aligned} \tag{V.34}$$

VI: We reprint the diagram here, but with labels:



(V.35)

Let K' be the $d \times d$ matrix such that

$$\langle j_a | K' | j_b \rangle = \sqrt{\sum_{j_c, j_d} \left(\langle j_a, j_c | \hat{\mathcal{Y}} | j_b, j_d \rangle \right)^2}. \tag{V.36}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

Then the contribution of the diagram is bounded by

$$\begin{aligned}
&\sum_{\substack{j_\alpha,j_\beta,j_\gamma,j_\delta\\ j_\xi,j_\chi,j_\phi,j_\psi}} \langle j_\xi,j_\beta|\hat{\mathcal{Y}}|j_\chi,j_\alpha\rangle\langle j_\phi,j_\alpha|\hat{\mathcal{Y}}|j_\psi,j_\beta\rangle \\
&\qquad\times \langle j_\psi,j_\delta|\hat{\mathcal{Y}}|j_\phi,j_\gamma\rangle\langle j_\chi,j_\gamma|\hat{\mathcal{Y}}|j_\xi,j_\delta\rangle \\
&\le \sum_{\substack{j_\xi,j_\chi\\ j_\phi,j_\psi}} \sqrt{\sum_{j_\alpha,j_\beta}\Big(\langle j_\xi,j_\beta|\hat{\mathcal{Y}}|j_\chi,j_\alpha\rangle\Big)^2} \\
&\qquad\times\sqrt{\sum_{j_\alpha,j_\beta}\Big(\langle j_\phi,j_\alpha|\hat{\mathcal{Y}}|j_\psi,j_\beta\rangle\Big)^2} \\
&\qquad\times\sqrt{\sum_{j_\gamma,j_\delta}\Big(\langle j_\psi,j_\delta|\hat{\mathcal{Y}}|j_\phi,j_\gamma\rangle\Big)^2} \\
&\qquad\times\sqrt{\sum_{j_\gamma,j_\delta}\Big(\langle j_\chi,j_\gamma|\hat{\mathcal{Y}}|j_\xi,j_\delta\rangle\Big)^2} \\
&\le \sum_{\substack{j_\xi,j_\chi\\ j_\phi,j_\psi}} \langle j_\xi|K'|j_\chi\rangle\langle j_\chi|K'|j_\xi\rangle\langle j_\phi|K'|j_\psi\rangle\langle j_\psi|K'|j_\phi\rangle \\
&\qquad= \operatorname{tr}(K'^2)^2 \\
&\qquad\le \|K'\|_F^4 \\
&\qquad= \left(\sum_{j_a,j_b}\sum_{j_c,j_d}\Big(\langle j_a,j_c|\hat{\mathcal{Y}}|j_b,j_d\rangle\Big)^2\right)^2 \\
&\qquad= \|\hat{\mathcal{Y}}\|_F^4 \\
&\qquad= d^4\|J(\mathcal{Y})\|_F^4. \qquad \text{(V.37)}
\end{aligned}$$

VII: We have a contribution of

$$\begin{aligned}
\operatorname{tr}(\hat{\mathcal{Y}}^2)\operatorname{tr}(\hat{\mathcal{Y}}^2) &\le \|\hat{\mathcal{Y}}\|_F^4 \\
&\le d^4\|J(\mathcal{Y})\|_F^4. \qquad \text{(V.38)}
\end{aligned}$$

Looking at all possible diagrams, we see that the total contribution of any diagram is less than

$$d^4\max\{\operatorname{tr}(J(\mathcal{Y}))^4, \|J(\mathcal{Y})\|_F^4\}. \qquad \text{(V.39)}$$

Hence this bounds the size of any term in Eq. (V.18). Each term in Eq. (V.18) is multiplied by a Weingarten function $W_g(d,4,\sigma\tau^{-1})$. The largest Weingarten term is

$$\begin{aligned}
&W_g(d,4,(1234)) = \\
&\frac{d^4-8d^2+6}{(d+3)(d+2)(d+1)(d-1)(d-2)(d-3)d^2} \\
&<\frac{2}{d^4} \qquad \text{(V.40)}
\end{aligned}$$

for $d>3$. There are $(4!)^2$ total terms in the sum. Putting it all together, we have

$$\mathbb{E}[\hat{S}^4] \le 2(4!)^2 d^4\max\{\operatorname{tr}(J(\mathcal{Y}))^4, \|J(\mathcal{Y})\|_F^4\} \qquad \text{(V.41)}$$

for $d>3$.

Now we combine Eqs. (V.12), (V.13), and (V.41), and the Payley-Zygmund inquality to obtain

$$\begin{aligned}
\mathbb{P}\left[\hat{S}\ge\xi\right] &\ge \mathbb{P}\left[\hat{S}^2\ge\xi^2\mathbb{E}[\hat{S}^2]\right] \\
&\ge \frac{(1-\xi^2)^2(\mathbb{E}\hat{S}^2)^2}{\mathbb{E}\hat{S}^4} \\
&\ge \frac{(1-\xi^2)^2\frac{1}{4}d^4\max\{\operatorname{tr}(J(\mathcal{F}))^4, \|J(\mathcal{F})\|_F^4\}}{2(4!)^2d^4\max\{\operatorname{tr}(J(\mathcal{F}))^4, \|J(\mathcal{F})\|_F^4\}} \\
&\ge \frac{(1-\xi^2)^2}{8(4!)^2}. \qquad \text{(V.42)}
\end{aligned}$$

Thus

$$\hat{Q}_\xi(\hat{E}) \ge \frac{(1-\xi^2)^2}{8(4!)^2}. \qquad \text{(V.43)}$$

D. Upper-bounding $\hat{W}_m(\hat{E})$

In this section, we will bound

$$\hat{W}_m(\hat{E}) = \mathbb{E}\sup_{Y\in\hat{E}}\operatorname{tr}(Y\hat{H}), \quad \hat{H} = \frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i J(\mathcal{C}_i)d^2, \qquad \text{(V.44)}$$

where the ε_i are Rademacher random variables, and the expectation is taken both over the ε_i and the choice of the unitaries C_i in the unitary 4-design. Because a unitary 4-design is also a unitary 2-design, a nearly identical argument to that used in Sec. IV-F holds, and we have

$$\hat{W}_m(\hat{E}) \le c_5\sqrt{\ln d}\cdot d, \qquad \text{(V.45)}$$

where $c_5>0$ is some numerical constant.

E. Final result

Combining equations (V.4), (V.7), (V.43) and (V.45), and setting $\xi=1/4$ and $t=\frac{1}{16}\cdot\frac{1}{8(4!)^2}\sqrt{m}$, we get that:

$$\begin{aligned}
&\inf_{\mathcal{U}}\lambda_{\min}(\hat{\mathcal{A}}; D(\hat{f}, J(\mathcal{U})d, 0)) \\
&\quad\ge \tfrac{1}{4}\sqrt{m}\tfrac{9}{16}\cdot\tfrac{1}{8(4!)^2} - 2c_5\sqrt{\ln d}\cdot d - \tfrac{1}{4}t \qquad \text{(V.46)}\\
&\quad= \tfrac{1}{4}\sqrt{m}\tfrac{1}{2}\cdot\tfrac{1}{8(4!)^2} - 2c_5\sqrt{\ln d}\cdot d.
\end{aligned}$$

Now we set $m\ge\left(64(4!)^2c_0\right)^2\cdot d^2\ln d$, which implies

$$\frac{1}{64(4!)^2c_0}\sqrt{m} \ge \sqrt{\ln d}\cdot d, \qquad \text{(V.47)}$$

hence

$$\begin{aligned}
&\inf_{\mathcal{U}}\lambda_{\min}(\hat{\mathcal{A}}; D(\hat{f}, J(\mathcal{U})d, 0)) \\
&\quad\ge \frac{\sqrt{m}}{64(4!)^2}\left(1-\frac{2c_5}{c_0}\right). \qquad \text{(V.48)}
\end{aligned}$$

This can be plugged into our approximate recovery bound (V.3). This finishes the proof of Theorem I.1. □

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

VI. Quantum Process Tomography

A. Motivation

Quantum process tomography is an important tool for the experimental development of large-scale quantum information processors. Process tomography is a means of obtaining complete knowledge of the dynamical evolution of a quantum system. This allows for accurate calibration of quantum gates, as well as characterization of qubit noise processes, such as dephasing, leakage, and cross-talk. These are the types of error processes that must be understood and ameliorated in order build scalable quantum computers with error rates below the fault-tolerance threshold.

Formally, a quantum process $\mathcal{E}$ is described by a completely-positive trace-preserving linear map $\mathcal{E} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$, which maps density matrices to density matrices. We want to learn $\mathcal{E}$ by applying it on known input states ρ, and measuring the output states $\mathcal{E}(\rho)$.

Process tomography is challenging for (at least) two reasons. First, it requires estimating a *large* number of parameters: for a system of n qubits, the Hilbert space has dimension $d = 2^n$, and a general quantum process has approximately $d^4 = 2^{4n}$ degrees of freedom. For example, to characterize the cross-talk between a pair of two-qubit gates acting on $n = 4$ qubits, using the most general approach, one must estimate $\sim 2^{16}$ parameters. To address this issue, several authors have proposed methods based on *compressed sensing*, which exploit sparse or low-rank structure in the unknown state or process, to reduce the number of measurements that must be performed [3], [5], [40]–[44].

The second challenge is that, in most real experimental setups, one must perform process tomography using state preparation and measurement devices that are *imperfect*. These devices introduce state preparation and measurement errors ("SPAM errors"), which limit the accuracy of process tomography. This limit is encountered in practice, and often produces non-physical estimates of processes [49]. Perhaps surprisingly, there are methods that are robust to SPAM errors, such as *randomized benchmarking* [45]–[49], and gate set tomography [37]–[39].

Here, we show how our phase retrieval techniques can address both of these challenges. We focus on the special case where we want to learn a *unitary* quantum process. This is a process $\mathcal{U} : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ that has the form

$$\mathcal{U} : \rho \mapsto U\rho U^\dagger, \tag{VI.1}$$

where $U \in \mathbb{C}^{d\times d}$ is a unitary matrix. This describes the dynamics of a closed quantum system, e.g., time evolution generated by some Hamiltonian, or the action of unitary gates in a quantum computer. Using our phase retrieval techniques, we devise methods for learning $\mathcal{U}$ that are both fast *and* robust to SPAM errors (at least in principle).

Whereas a generic quantum process would have $\sim d^4$ degrees of freedom, $\mathcal{U}$ has only $\sim d^2$ degrees of freedom. Our phase retrieval techniques work by estimating $m = O(d^2 \text{poly}(\log d))$ parameters of $\mathcal{U}$, hence they achieve a quadratic speedup over conventional tomography. Furthermore, we show how these measurements can be implemented using commonly-available quantum operations. In particular, they can be performed using randomized benchmarking protocols, which are robust to SPAM errors.

B. Measurement Techniques

In order to learn the unknown process $\mathcal{U}$, we would like to measure the quantities $|\text{tr}(C_i^\dagger U)|^2$, where $C_1, \dots, C_m$ are chosen at random from a unitary 4- or 2-design in $\mathbb{C}^{d\times d}$. Then we can apply the PhaseLift algorithm to reconstruct U.

We remark that these measurements are quite different from the Pauli measurements that were used in earlier works on compressed sensing for quantum tomography [3]–[5]. In those earlier works, one would estimate the Pauli expectation values of the Jamiolkowski state $J(\mathcal{U})$. This led to measurements of the form $\text{tr}((P_1 \otimes P_2)J(\mathcal{U})) = \frac{1}{d}\text{tr}(P_1 U P_2 U^\dagger)$, where P_1 and P_2 were multi-qubit Pauli operators.

There are (at least) two ways of measuring the quantity $|\text{tr}(C^\dagger U)|^2$, for some prescribed unitary $C \in \mathbb{C}^{d\times d}$. The first way works by estimating the Choi-Jamiolkowski state of U. This method is straightforward, and fairly efficient, but it requires reliable and error-free state preparations and measurements.

To apply this method, recall that any quantum process $\mathcal{E}$ can be equivalently described by its Choi-Jamiolkowski state,

$$J(\mathcal{E}) = (\mathcal{E} \otimes \mathcal{I})(|\Phi^+\rangle\langle\Phi^+|) \in \mathbb{C}^{d^2\times d^2}, \tag{VI.2}$$

which is obtained by applying $\mathcal{E}$ to one half of the maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1} |i\rangle \otimes |i\rangle \in \mathbb{C}^{d^2}$. In the case of a unitary process $\mathcal{U}$, this is equivalent to the "lifted" representation of U used in PhaseLift:

$$\begin{aligned} J(\mathcal{U}) &= (U \otimes I)|\Phi^+\rangle\langle\Phi^+|(U^\dagger \otimes I) \\ &= \frac{1}{d}\text{vec}(U)\text{vec}(U)^\dagger. \end{aligned} \tag{VI.3}$$

In particular, this implies that

$$\begin{aligned} |\text{tr}(C^\dagger U)|^2 &= d^2\,\text{tr}(J(\mathcal{C})J(\mathcal{U})) \\ &= d^2\left|\langle\Phi^+|(C^\dagger \otimes I)(U \otimes I)|\Phi^+\rangle\right|^2. \end{aligned} \tag{VI.4}$$

Thus, the quantity $|\text{tr}(C^\dagger U)|^2$ can be estimated via the following procedure: (1) prepare a maximally entangled state $|\Phi^+\rangle$ on two registers; (2) apply the unknown unitary operation U (on the first register); (3) apply the prescribed unitary operation $C^\dagger$ (on the first register); (4) measure both registers in the Bell bases; (5) repeat the above steps, and count the number of times that the outcome $|\Phi^+\rangle$ is observed.

The second way of estimating $|\text{tr}(C^\dagger U)|^2$ makes use of a more sophisticated technique known as randomized benchmarking tomography [48], [49]. This method is robust to SPAM errors, but it is also quite resource-intensive. In addition, this method requires that the measurement matrix C be chosen from some group that has a computationally efficient description, such as the Clifford group [26], [27].

In randomized benchmarking tomography, one implements a sequence that alternates between applying a random Clifford

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

and applying the unknown unitary map $\mathcal{U}$. This alternation repeats L times, and then is followed by a single recovery Clifford operation. For example, the recovery operation could be the Clifford that inverts the action of the L randomly applied Cliffords in the sequence, so the total effect of the $L+1$ applied Cliffords (ignoring the applications of $\mathcal{U}$) would be the identity.

This protocol, when repeated many times, and for many different lengths L, produces a measurement signature of a decaying exponential in L, where the rate of decay depends only on $\mathrm{tr}(\mathcal{U})$. By choosing different recovery operations, one can cause the rate of decay to depend on $\mathrm{tr}(\mathcal{U}\mathcal{C}^\dagger)$ for any Clifford C. (This analysis assumes that Clifford operations can be implemented perfectly. If the Clifford operations contain errors, this procedure still works, and it produces a characterization of the process $\mathcal{U} \circ \Lambda$, where Λ is the average Clifford error.)

To summarize, randomized benchmarking tomography allows us to estimate quantities of the form $\mathrm{tr}(\mathcal{U}\mathcal{C}^\dagger)$, where we can choose $\mathcal{C} : \rho \mapsto C\rho C^\dagger$ to be any Clifford operation. We can rewrite this as

$$\mathrm{tr}(\mathcal{U}\mathcal{C}^\dagger) = d^2\,\mathrm{tr}(J(\mathcal{U})J(\mathcal{C})) = |\mathrm{tr}(C^\dagger U)|^2. \qquad \text{(VI.5)}$$

C. Numerical Simulation

We ran numerical simulations to assess the performance of the PhaseLift algorithm for reconstructing an unknown unitary operation $\mathcal{U} : \rho \mapsto U\rho U^\dagger$ (where $U \in \mathbb{C}^{d\times d}$), using random Clifford measurements $C_1, \ldots, C_m \in \mathbb{C}^{d\times d}$. In our simulations, we chose U at random, by sampling from the Haar distribution on the unitary group. We then sampled the C_i from the uniform distribution on the Clifford group. We compared this with a second scenario, where the C_i were sampled from the Haar distribution on the unitary group, since this is the "best" measurement ensemble for phase retrieval.

We simulated the measurement procedure (I.6) on $\mathcal{U}$, in the noiseless case ($\varepsilon = 0$). Then we solved the PhaseLift convex program (I.7), again in the noiseless case ($\eta = 0$), in order to obtain an estimate $\mathcal{U}'$ of $\mathcal{U}$. Here we omitted the non-spikiness constraints (I.14), in order to make the convex program easier to solve. Finally, we computed the reconstruction error in the Frobenius norm, $\|J(\mathcal{U}') - J(\mathcal{U})\|_F$.

The results are shown in Figure 1. These results suggest that phase retrieval using random Clifford measurements performs nearly as well as phase retrieval using Haar-random unitary measurements. In both cases, the number of measurements scales as $m \approx Cd^2$, where $C \approx 4.8$. This suggests that, although random Clifford operations are not a unitary 4-design, they are "close enough" to ensure accurate reconstruction of unitary matrices via phase retrieval.

Acknowledgments

The authors thank Richard Kueng and David Gross for helpful discussions about this problem; Easwar Magesan for sharing his Clifford sampling code; Aram Harrow and Steve Flammia for clarifications regarding the results in [13]; and several anonymous reviewers for suggestions which improved this paper. Contributions to this work by NIST, an agency of the US government, are not subject to US copyright.

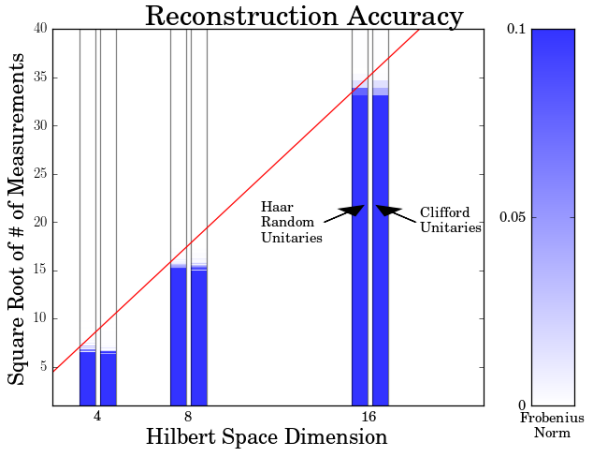


Fig. 1. Phase retrieval of an unknown unitary matrix $U \in \mathbb{C}^{d\times d}$, using measurement matrices $C_1, \ldots, C_m \in \mathbb{C}^{d\times d}$ which are chosen to be either Haar random unitaries or random Cliffords. The x-axis shows the dimension d, and the y-axis shows the square root $\sqrt{m}$ of the number of measurements. For each choice of dimension $d \in \{2^2, 2^3, 2^4\}$, the two vertical bars correspond to the two cases where the measurement matrices are either Haar random unitaries or random Cliffords. The color scale shows the error of the reconstructed matrix U', measured in Frobenius norm; white denotes near-perfect reconstruction. Each data point was averaged over 10 runs, where each run used fresh random samples of U and $C_1, \ldots, C_m$. The red line shows the expected linear scaling, $\sqrt{m} = \alpha d$, where $\alpha = 2.2$ was chosen to fit the data. Somewhat surprisingly, random Clifford measurements perform nearly as well as measurements using Haar-random unitaries.

References

[1] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Optical Society of America A* 7(3), pp.394-411 (1990).

[2] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao and M. Segev, "Phase Retrieval with Application to Optical Imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87-109, May 2015.

[3] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[4] Y.-K. Liu. Universal low-rank matrix recovery from pauli measurements. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011*, pages 1638–1646, 2011.

[5] S. T Flammia, D. Gross, Y.-K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

[6] E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commum. Pure and Applied Math.*, 66(8):1241–1274, 2013.

[7] D. Gross, F. Krahmer, and R. Kueng. A partial derandomization of phaselift using spherical designs. *J. Fourier Analysis and Applications*, 21(2):229–266, 2015.

[8] E. J. Candes and X. Li. Solving quadratic equations via phaselift when there are about as many equations as unknowns. *Found. Comput. Math.*, 14(5):1017–1026, 2014.

[9] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Comput. Harmonic Analysis* 42(1), pp.88-116, January 2017.

[10] R. Kueng, H. Zhu, and D. Gross. Low rank matrix recovery from Clifford orbits. arXiv:1610.08070.

[11] R. Balan, B. G. Bodmann, P. G. Casazza, and D. Edidin. Painless reconstruction from magnitudes of frame coefficients. *J. Fourier Anal. Appl.*, vol. 15, no. 4, pp. 488–501, 2009.

[12] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase Retrieval via Matrix Completion," *SIAM J. Imaging Sci.*, 6(1), pp.199–225 (2013).

[13] C.B. Mendl and M.M. Wolf, "Unital Quantum Channels - Convex Structure and Revivals of Birkhoff's Theorem," *Commun. Math. Phys.* 289, 1057-1096 (2009).

[14] A. Ambainis and A. D. Smith, "Small Pseudo-random Families of Matrices: Derandomizing Approximate Quantum Encryption," *Proc. APPROX-RANDOM 2004*, pp.249-260 (2004).

[15] J. Emerson et al, "Symmetrised Characterisation of Noisy Quantum Processes," *Science* 317, 1893-1896 (2007).

[16] A. Ambainis and J. Emerson, "Quantum t-designs: t-wise independence in the quantum world," *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pp.129-140 (2007).

[17] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009.

[18] O. Szehr, F. Dupuis, M. Tomamichel and R. Renner, "Decoupling with unitary approximate two-designs," *New J. Phys.* 15 (2013) 053022.

[19] D. Gross, K. Audenaert and J. Eisert, "Evenly distributed unitaries: on the structure of unitary designs," *J. Math. Phys.* 48, 052104 (2007).

[20] R. A. Low, *Pseudo-randomness and Learning in Quantum Computation*, PhD Thesis, University of Bristol, UK (2010).

[21] F. G. S. L. Brandao, A. W. Harrow and M. Horodecki, "Local random quantum circuits are approximate polynomial-designs," Arxiv:1208.0692.

[22] R. Cleve, D. Leung, L. Liu, and C. Wang. Near-linear constructions of exact unitary 2-designs. *Quantum Information and Computation*, 16:0721, 2015.

[23] R. Kueng and D. Gross. Qubit stabilizer states are complex projective 3-designs. arXiv:1510.02767, 2015.

[24] Z. Webb. The clifford group forms a unitary 3-design. arXiv:1510.02769, 2015.

[25] H. Zhu. Multiqubit clifford groups are unitary 3-designs. arXiv:1510.02619, 2015.

[26] D. Gottesman. Stabilizer codes and quantum error correction. *arXiv preprint quant-ph/9705052*, 1997.

[27] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal. Quantum data hiding. *Information Theory, IEEE Transactions on*, 48(3):580–598, 2002.

[28] V. Koltchinskii and S. Mendelson. Bounding the Smallest Singular Value of a Random Matrix Without Concentration. *Int Math Res Notices* (2015) 2015 (23): 12991-13008.

[29] S. Mendelson. Learning without Concentration. *Journal of the ACM* 62(3):21, 2015.

[30] J. A. Tropp. Convex recovery of a structured signal from independent random linear measurements. In *Sampling Theory, a Renaissance*, G.E. Pfander (ed.), Springer (2015), pp.67-101.

[31] B. Collins. Moments and cumulants of polynomial random variables on unitary groups, the itzykson-zuber integral and free probability. *Int. Math. Res. Notes*, 17, 2003.

[32] B. Collins and P. Sniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.*, 264, 2006.

[33] A. J. Scott. Optimizing quantum process tomography with unitary 2-designs. *Journal of Physics A: Mathematical and Theoretical*, 41(5):055308, 2008.

[34] E. J. Candes and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[35] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, March 2011.

[36] S. Negahban and M. J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.*, 13:1665–1697, May 2012.

[37] S. T. Merkel, et al. Self-consistent quantum process tomography. *Phys. Rev. A*, 87:062119, Jun 2013.

[38] C. Stark. Self-consistent tomography of the state-measurement gram matrix. *Phys. Rev. A*, 89:052109, May 2014.

[39] D. Greenbaum. Introduction to quantum gate set tomography. ArXiv:1509.02921, 2015.

[40] M. Cramer et al. Efficient quantum state tomography. *Nature Commun.*, 1(149), 2010.

[41] A. Shabani, et al. Efficient measurement of quantum dynamics via compressive sensing. *Phys. Rev. Lett.*, 106:100401, Mar 2011.

[42] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz. Estimation of many-body quantum hamiltonians via compressive sensing. *Phys. Rev. A*, 84:012107, Jul 2011.

[43] Charles H. Baldwin, Amir Kalev, and Ivan H. Deutsch. Quantum process tomography of unitary and near-unitary maps. *Phys. Rev. A*, 90:012110, Jul 2014.

[44] A. Kalev, R. L. Kosut, and I. H. Deutsch. Informationally complete measurements from compressed sensing methodology. ArXiv:1502.00536, 2015.

[45] E. Knill, et al. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, Jan 2008.

[46] E. Magesan, J. M. Gambetta, and J. Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:180504, May 2011.

[47] Easwar Magesan, et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.*, 109:080505, Aug 2012.

[48] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki. Robust extraction of tomographic information via randomized benchmarking. *Phys. Rev. X*, 4:011050, Mar 2014.

[49] Blake R Johnson, et al. Demonstration of robust quantum gate tomography via randomized benchmarking. *New Journal of Physics*, 17(11):113019, 2015.

[50] H. Zhu, R. Kueng, M. Grassl, and D. Gross. The Clifford group fails gracefully to be a unitary 4-design. arXiv:1609.08172.

[51] J. Helsen, J. J. Wallman, and S. Wehner. Representations of the multiqubit Clifford group. arXiv:1609.08188.

[52] E. Candes, X. Li and M. Soltanolkotabi, "Phase Retrieval via Wirtinger Flow: Theory and Algorithms," *IEEE Trans. Info. Theory*, Vol. 64 (4), Feb. 2015.

[53] J. Matousek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[54] S. Popescu, A. J. Short, and A. Winter. Entanglement and the foundations of statistical mechanics. *Nature Physics*, 2(11):754–758, 2006.

[55] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[56] R. A. Low. Large deviation bounds for k-designs. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465(2111):3289–3308, 2009.

[57] R. Vershynin. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

APPENDIX

Given a completely positive and trace preserving quantum operation $\mathcal{F} : \mathbb{C}^{d\times d} \to C^{d\times d}$, there are several useful representation of $\mathcal{F}$. One is the Choi-Jamiolkowski representation $J(\mathcal{F}) \in \mathbb{C}^{d^2\times d^2}$,

$$J(\mathcal{F}) = (\mathcal{F}\otimes\mathcal{I})(|\Phi^+\rangle\langle\Phi^+|), \tag{A.1}$$

which is obtained by applying $\mathcal{F}$ to one half of the maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}|i\rangle\otimes|i\rangle \in \mathbb{C}^{d^2}$. Another is the Liouville representation $\mathcal{F}^L \in \mathbb{C}^{d^2\times d^2}$,

$$\begin{aligned}\langle kl|\mathcal{F}^L|ij\rangle &= \langle k|\mathcal{F}(|i\rangle\langle j|)|l\rangle \\ &= d\langle ki|J(\mathcal{F})|lj\rangle,\end{aligned} \tag{A.2}$$

where $|i\rangle$, $|j\rangle$, $|l\rangle$ and $|k\rangle$ are any standard basis states.

All representations are completely equivalent, and it is a simple exercise to convert between them. However, certain representations make it easier to check for properties like complete positivity.

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

For example from the complete positivity constraint, we have that $J(\mathcal{X})$ is Hermitian, so

$$\langle ki|J(\mathcal{X})|lj\rangle = \langle lj|J(\mathcal{X})^*|ki\rangle. \tag{A.3}$$

Converting this into Liouville representation, we have

$$\langle kl|\mathcal{X}^L|ij\rangle = \langle lk|(\mathcal{X}^L)^*|ji\rangle. \tag{A.4}$$

Using this fact, we can show that for completely positive superoperators $\mathcal{F}$ and $\mathcal{K}$,

$$\operatorname{tr}(J(\mathcal{F})J(\mathcal{K})) = \frac{1}{d^2}\operatorname{tr}(\mathcal{F}^L(\mathcal{K}^L)^\dagger). \tag{A.5}$$

To see this, we calculate

$$\begin{aligned}\operatorname{tr}(J(\mathcal{F})J(\mathcal{K})) =& \frac{1}{d^2}\operatorname{tr}\left(\sum_{ij}\mathcal{F}(|i\rangle\langle j|)\mathcal{K}(|j\rangle\langle i|)\right)\\ =& \frac{1}{d^2}\sum_{ijkl}\langle k|\mathcal{F}(|i\rangle\langle j|)|l\rangle\langle l|\mathcal{K}(|j\rangle\langle i|)|k\rangle\\ =& \frac{1}{d^2}\sum_{ijkl}\langle kl|\mathcal{F}^L|ij\rangle\langle kl|(\mathcal{K}^L)^*|ij\rangle\\ =& \frac{1}{d^2}\operatorname{tr}(\mathcal{F}^L(\mathcal{K}^L)^\dagger).\end{aligned} \tag{A.6}$$

Additionally, we note that for unitary maps $\mathcal{U}:\rho\mapsto U\rho U^\dagger$, where U is a unitary operation, the corresponding Liouville representation takes the form

$$\mathcal{U}^L = U\otimes U^*, \tag{A.7}$$

because

$$\mathcal{U}(|i\rangle\langle j|) = U|i\rangle\langle j|U^\dagger \tag{A.8}$$

Using Eq. (A.5) and Eq. (A.7), we have that for a map $\mathcal{U}$ representing a unitary U and a map $\mathcal{C}$ representing a unitary C, that

$$\operatorname{tr}(J(\mathcal{U})J(\mathcal{C})) = (1/d^2)|\operatorname{tr}(U^\dagger C)|^2. \tag{A.9}$$

Liu, Yi-Kai; Kimmel, Shelby.
"Phase Retrieval Using Unitary 2-Designs."
Paper presented at 2017 International Conference on Sampling Theory and Applications (SampTA 2017), Tallinn, Estonia. July 3, 2017 - July 7, 2017.

Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements

Thomas A. Germer

Sensor Science Division
National Institute of Standards and Technology
Gaithersburg, MD 20899 USA

ABSTRACT

We report on Mueller matrix bidirectional reflectance distribution function (BRDF) measurements of a sintered polytetrafluoroethylene (PTFE) sample over the scattering hemisphere for incident angles from 0° to 75° and for four wavelengths from 351 nm to 1064 nm. The data are fit to a radiative transfer equation solution for a diffuse medium, letting parameters describing the single scattering phase function be adjusted.

Keywords: BRDF, Mueller matrix, PTFE, radiative transfer, reflectance standards

1. INTRODUCTION

Having high reflectance from the near ultraviolet to the near infrared and being relatively Lambertian, pressed polytetrafluoroethylene (PTFE) powder was developed as a reference standard for diffuse reflectance and is used for constructing integrating spheres for irradiance averaging devices and uniform radiance sources.[1–3] A sintered, machinable version of the material is sold under the trade names Labsphere Spectralon, Avian Technologies Fluorilon-99W, SphereOptics Zenith, Lake Photonics Spectralex, Gigahertz-Optik OP.DI.MA, and others.* Much of its reflectance properties are well documented. Many studies considered the like- and crossed-polarized reflectance properties in the plane of incidence.[4–9] Mueller matrix properties in the plane of incidence are reported in Refs. 11–15. A small number of measurements are reported out of the plane of incidence,[16,17] while others considered the wavelength dependence of the optical properties.[1–3,10,11]

Recently, we performed and published a comprehensive Mueller matrix bidirectional reflectance distribution function (BRDF) measurement for a sintered PTFE sample, covering the hemisphere with over 300 sampled points at each of six incident angles ranging from normal incidence to 75° and for four wavelengths (351 nm, 532 nm, 633 nm, and 1064 nm).[18] Expanding upon the work of Koenderink and van Doorn (KvD),[19] the data were fit to a phenomenological description of the Mueller matrix BRDF that, by use of symmetric products of Zernike polynomials, is complete for smooth functions and obeys reciprocity. That description allows for an on-demand evaluation of the Mueller matrix BRDF for any scattering geometry. The phenomenological description of the scattering function for sintered PTFE fully describes the reflective scattering properties of a thick sample, but it lacks any physical insight into how the material might behave if, for example, an absorber were added, the grain size changed, or the surface smoothened.

In this work, we consider a radiative transfer equation (RTE) model with a generalized polarimetric single scattering phase function. By fitting the data taken in Ref. 18, we find that we can determine the single scattering phase function for this material.

In Sec. 2, we describe the measurements and present the results. In Sec. 3, we outline the radiative transfer equation, its solution, and how we generalize it to include polarization. In Sec. 4, we describe the results of fitting the data to the polarized radiative transfer equation solution.

Corresponding author: thomas.germer@nist.gov, Telephone: 1 301 975 2876

*Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is neither intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

2. MEASUREMENT

Measurements were carried out using the laser-based goniometric optical scatter instrument (GOSI) at the National Institute of Standards and Technology (NIST).[20] GOSI measures the Mueller matrix BRDF,[21] $\mathbf{f}_{\mathrm{r}}(\theta_{\mathrm{i}}, \phi_{\mathrm{i}}, \theta_{\mathrm{r}}, \phi_{\mathrm{r}})$, for any pair of incident (defined by polar angle θ_{i} and azimuthal angle ϕ_{i}) and scattering (defined by polar angle θ_{r} and azimuthal angle ϕ_{r}) directions at a number of laser wavelengths. Four laser sources were used for these measurements: a continuous wave (cw) Nd-doped yttrium-aluminum-garnet (Nd:YAG) laser operating at 1064 nm, a cw HeNe laser operating at 633 nm, a frequency-doubled cw Nd:YAG laser operating at 532 nm, and a frequency-doubled mode-locked Ti:sapphire laser operating at 351 nm (fundamental at 702 nm). The solid angle subtended by the receiver was 1.114×10^{-4} sr. The sample was an 8 mm thick, 50 mm diameter coupon of sintered PTFE obtained commercially. The material was unsupported, and the back was left open during the measurements. Data were taken for six incident angles (0°, 15°, 30°, 45°, 60°, and 75°), measuring the scatter on a uniform grid in direction cosine space, with direction cosine spacing of 0.1, so that there were 304 scattering directions measured for each incident angle. Points (6 total) were removed from each wavelength's dataset if the receiver was blocking the incident beam.

Mueller matrix measurements were performed using a rotating-retarder polarization state generator and a rotating-retarder polarization state analyzer. Sixteen measurements in a 4×4 grid were performed at each scattering geometry. The reduction matrix converting the sixteen measurements to the Mueller matrix was determined using the method of Compain *et al.*[22] This method uses a high quality polarizer (in this case a Glan-Taylor polarizer) in transmission and a diattenuator-retarder (in this case a silicon wafer with a nominal 1000 nm thermally-grown SiO_2 layer) in reflection. While the method calls for one full Mueller matrix measurement from each of the two references and one without any sample present, we perform measurements of the polarizer in transmission for 18 different rotations (0° to 170°), measurements of the diattenuator-retarder for 18 angles of incidence (30° to 75°), and 18 measurements without any reference at all. Thus, there was significant redundancy in our calibration process that allows us to determine various figures of merit for the accuracy of subsequent measurements. For example, the standard deviations of all the reference measurements across all elements of the normalized Mueller matrix were 0.0043 (for 351 nm), 0.0014 (for 532 nm), 0.0023 (for 633 nm), and 0.0052 (for 1064 nm). We believe we can use these standard deviations as standard uncertainties in subsequent measurements.

We choose to express the Mueller matrix BRDF in a polarization basis that avoids singularities in the incident or scattering hemispheres. In an s and p basis, defined as perpendicular and parallel, respectively, to the plane containing the direction of propagation and the surface normal (with different planes for incident and scattering directions), there exists a singularity along the surface normal. In a $\perp$ and $\parallel$ basis, defined as perpendicular and parallel, respectively, to the plane containing both the incident and scattering directions of propagation, there exists a singularity in the retroreflection direction. Therefore, we chose an x and y basis set, defined as the same as p and s, respectively, when $\phi_{\mathrm{i}} = 0$ and $\phi_{\mathrm{r}} = 0$, but which rotate by ϕ_{i} and ϕ_{r}, respectively, to remove the singularity existing in the s-p basis at the surface normal. We consider y, x, and the direction of propagation to form a right-handed triplet. The x-y basis is the natural basis in a back-focal-plane microscope scatterometer. Here, we define the four elements of Stokes vectors by

$$\begin{aligned} S_0 &= I_y + I_x, \\ S_1 &= I_y - I_x, \\ S_2 &= I_{y+x} - I_{y-x}, \\ S_3 &= I_{\mathrm{l}} - I_{\mathrm{r}}, \end{aligned} \tag{1}$$

where I_y, I_x, etc. are intensity-like quantities (e.g., irradiance or radiance). I_{l} and I_{r} are intensities for left and right circular polarizations, where we use the convention that right-handed polarization refers to the electric field rotating clockwise when viewing into the beam.

The actual Mueller matrix BRDF data have larger uncertainties than those estimated by the calibration process, due to random errors imposed by laser speckle and the lower signal levels of the scatter measurement; these can be estimated by observing the point-to-point variations in the data. The apparent noise is higher for the 1064 nm results, partly because of the longer wavelength, but also since the focussing of the system had been changed so that the spot size at the sample was smaller.

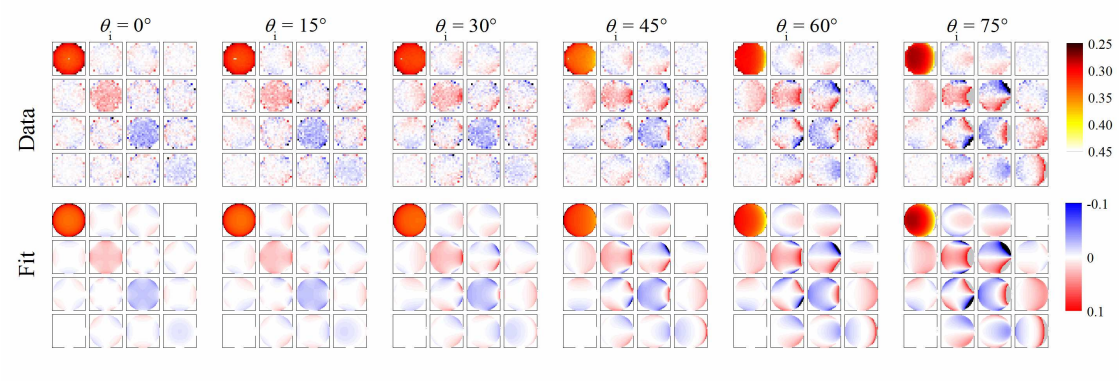


Figure 1. The Mueller matrix BRDF at 351 nm shown in direction cosine space (vertical and horizontal scales −1 to 1). Each column corresponds to an incident angle. The top row shows the measured data, while the second row shows the fit to the radiative transfer equation model. The 00 element is on a scale from (black) 0.25 sr^{-1} to (white) 0.45 sr^{-1}, while the other elements are shown normalized on a scale from (blue) −0.1 to (red) 0.1. Some regions are saturated on this scale.

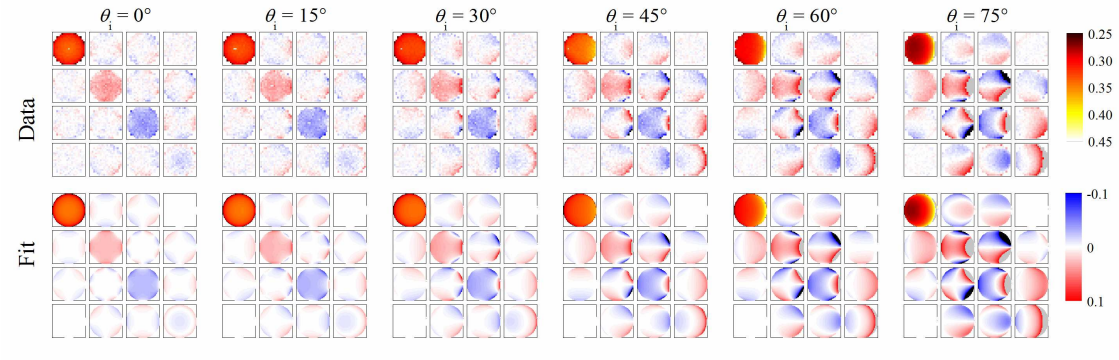


Figure 2. The Mueller matrix BRDF at 532 nm shown in direction cosine space (vertical and horizontal scales −1 to 1). Each column corresponds to an incident angle. The top row shows the measured data, while the second row shows the fit to the radiative transfer equation model. The 00 element is on a scale from (black) 0.25 sr^{-1} to (white) 0.45 sr^{-1}, while the other elements are shown normalized on a scale from (blue) −0.1 to (red) 0.1. Some regions are saturated on this scale.

The results of the measurements are shown in the upper rows of Figs. 1–4 for 351 nm, 532 nm, 633 nm, and 1064 nm, respectively. The results are shown in direction cosine space, so that the radial coordinate represents $\sin\theta_\mathrm{r}$, and the plane of incidence is a horizontal line through the center of each circle. The lower rows of Figs. 1–4 show the best fit to the data as described later in this paper.

3. THEORY

3.1 Scalar radiative transfer equation

The scalar form of the radiative transfer equation for a stratified, possibly anisotropic, medium, ignoring emission, is

$$\cos\theta \frac{\mathrm{d}L(z,\Omega)}{\mathrm{d}z} = -A(z,\Omega)L(z,\Omega) - \int_{4\pi} G(z,\Omega',\Omega)L(z,\Omega)\,\mathrm{d}\Omega' + \int_{4\pi} G(z,\Omega,\Omega')L(z,\Omega')\,\mathrm{d}\Omega', \tag{2}$$

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

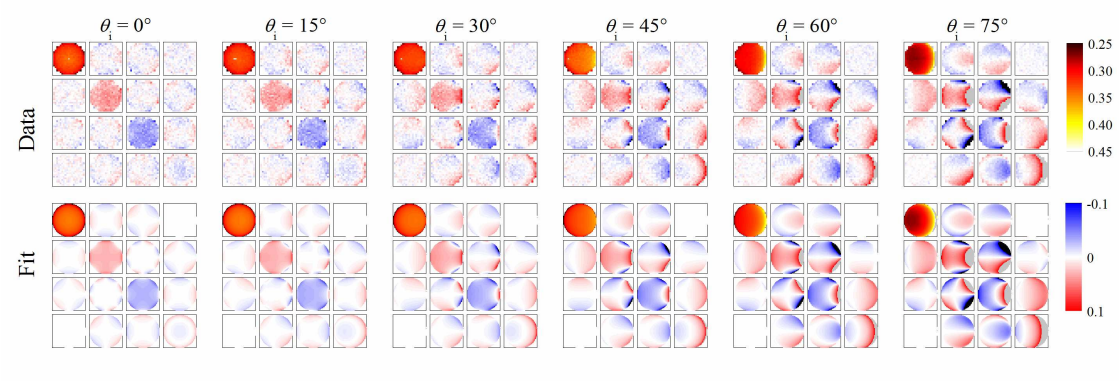


Figure 3. The Mueller matrix BRDF at 633 nm shown in direction cosine space (vertical and horizontal scales −1 to 1). Each column corresponds to an incident angle. The top row shows the measured data, while the second row shows the fit to the radiative transfer equation. The 00 element is on a scale from (black) 0.25 sr^{-1} to (white) 0.45 sr^{-1}, while the other elements are shown normalized on a scale from (blue) −0.1 to (red) 0.1. Some regions are saturated on this scale.

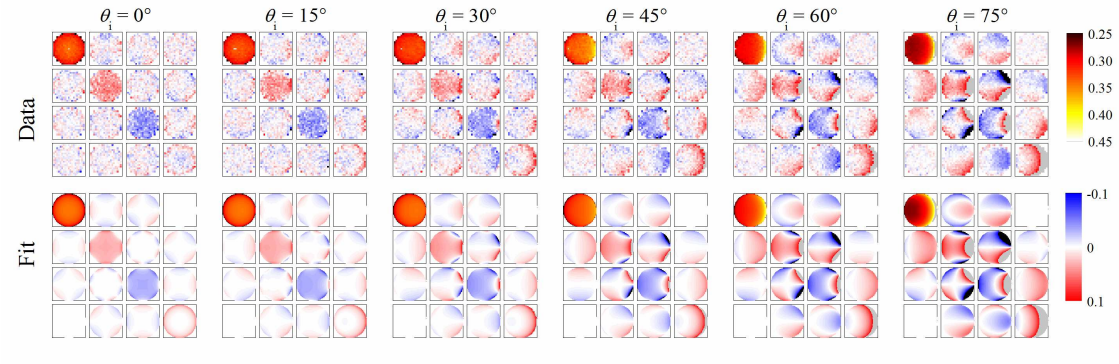


Figure 4. The Mueller matrix BRDF at 1064 nm shown in direction cosine space (vertical and horizontal scales −1 to 1). Each column corresponds to an incident angle. The top row shows the measured data, while the second row shows the fit to the radiative transfer equation model. The 00 element is on a scale from (black) 0.25 sr^{-1} to (white) 0.45 sr^{-1}, while the other elements are shown normalized on a scale from (blue) −0.1 to (red) 0.1. Some regions are saturated on this scale.

where $L(z,\Omega)$ is the radiance at depth z propagating in direction Ω, $A(z,\Omega)$ is the absorption coefficient at depth z for light propagating in direction Ω, $G(z,\Omega,\Omega')$ is the differential scattering coefficient at depth z for light propagating from direction Ω' and scattering to direction Ω, and θ is the polar angle of the direction of propagation. On the right-hand side of Eq. (2), the first term describes the absorption, the second term describes the losses from scattering away from the direction Ω, and the third term describes gains from scattering from other directions into direction Ω.

We can divide Eq. (2) by $\cos\theta$, separate each integral into upward and downward propagating hemispheres, change the integration variables from $\mathrm{d}\Omega' = \sin\theta'\ \mathrm{d}\theta'\mathrm{d}\phi'$ to $\sec\theta'\ \mathrm{d}x'\mathrm{d}y'$ (where $x' = \sin\theta'\cos\phi'$ and $y' = \sin\theta'\sin\phi'$ are the direction cosines), separate the directions into those propagating upward and downward, and

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

approximate the integrals by sums. Thus, for small Δz,

$$\Delta L_j^+ = \sum_k (T_{jk}^+ - \delta_{jk}) L_k^+ + \sum_k R_{jk}^+ L_k^-, \tag{3a}$$

$$\Delta L_j^- = \sum_k (T_{jk}^- - \delta_{jk}) L_k^- + \sum_k R_{jk}^- L_k^+, \tag{3b}$$

where

$$T_{jk}^\pm = \delta_{jk} - \frac{A(z,\Omega_j^\pm)}{\cos\theta_j}\Delta z\ \delta_{jk} - \frac{G(z,\Omega_k^\pm,\Omega_j^\pm)}{\cos\theta_j\cos\theta_k}\Delta z\ w_k - \frac{G(z,\Omega_k^\pm,\Omega_j^\mp)}{\cos\theta_j\cos\theta_k}\Delta z\ w_k + \frac{G(z,\Omega_j^\pm,\Omega_k^\pm)}{\cos\theta_j\cos\theta_k}\Delta z\ w_k, \tag{4a}$$

$$R_{jk}^\pm = \frac{G(z,\Omega_j^\pm,\Omega_k^\mp)}{\cos\theta_j\cos\theta_k}\Delta z\ w_k. \tag{4b}$$

L_k^- is the radiance in the downward direction Ω_k^-, while L_k^+ is radiance in the upward direction Ω_k^+. The directions and weights, w_k, are chosen to approximate the integral using Gaussian cubature on the unit circle using Zernike polynomials as the orthogonal polynomials.[23] The solution to Eqs. (3)–(4) can be found by using the adding-doubling method.[24] That is, if we create matrices $\mathbf{T}^\pm$ and $\mathbf{R}^\pm$ whose elements are $T_{jk}^\pm$ and $R_{jk}^\pm$, respectively, we can show that two layers (with $\mathbf{T}_1^\pm$ and $\mathbf{R}_1^\pm$ above $\mathbf{T}_2^\pm$ and $\mathbf{R}_2^\pm$ below) can be combined so that the net matrices are

$$\mathbf{R}_{12}^+ = \mathbf{R}_2^+ + \mathbf{T}_2^-(\mathbf{1} - \mathbf{R}_1^+\mathbf{R}_2^-)^{-1}\mathbf{R}_1^+\mathbf{T}_2^+, \tag{5a}$$

$$\mathbf{R}_{12}^- = \mathbf{R}_1^- + \mathbf{T}_1^+(\mathbf{1} - \mathbf{R}_2^-\mathbf{R}_1^+)^{-1}\mathbf{R}_2^-\mathbf{T}_1^-, \tag{5b}$$

$$\mathbf{T}_{12}^+ = \mathbf{T}_1^+(\mathbf{1} - \mathbf{R}_2^-\mathbf{R}_1^+)^{-1}\mathbf{T}_2^+, \tag{5c}$$

$$\mathbf{T}_{12}^- = \mathbf{T}_2^-(\mathbf{1} - \mathbf{R}_1^+\mathbf{R}_2^-)^{-1}\mathbf{T}_1^-. \tag{5d}$$

The reflection and transmission properties of a finite thickness of homogeneous material can thus be obtained by starting with a very thin layer and applying Eqs. (5) numerous times until the desired thickness is achieved.

Although the surface of the sintered PTFE is very rough, we designed the code to also handle the possibility of the top and bottom surfaces being smooth and exhibiting Fresnel reflection and refraction. This treatment is performed in two steps. First, we create a layer at each interface that exhibits the reflectances ($\mathbf{R}_\mathrm{t}^+$ for the top interface and $\mathbf{R}_\mathrm{b}^-$ for the bottom interface) predicted for a specific index of refraction using Fresnel theory and unit transmittance. We apply the adding-double analysis above, treating each interface as a new layer. All of the directions of propagation ($\Omega_j^\pm$) are treated as those inside the medium. In the second step, we apply Snell's law to relate the external angles of propagation with those internal to the layer and apply the Fresnel transmittance of the interface to determine the radiances outside the material.

The net matrices $\mathbf{R}_\mathrm{net}^\pm$ and $\mathbf{T}_\mathrm{net}^\pm$ define the relationship between the incident and scattered radiance. Since irradiance $E_k = L_k \cos\theta_k\ \mathrm{d}\Omega_k \to L_k w_k$, the BRDF is given by

$$f_\mathrm{r}(\Omega_k, \Omega_j) = [\mathbf{R}_\mathrm{net}^-]_{jk}/w_k. \tag{6}$$

3.2 Polarized radiative transfer equation

To generalize the radiative transfer solution above to include polarization, we can treat $L_j^\pm$ as four-element Stokes vectors and the absorption coefficients A and the diffential scattering coefficients G as 4×4-element Mueller matrices. However, we must consider two issues: we must ensure that the coordinate systems in which the Stokes vectors and Mueller matrices are expressed are consistent, and we need to treat the polarization-dependent loss from scattering away from Ω correctly.

It is natural to express the differential scattering coefficient using a local basis defined with respect to the plane containing Ω and Ω'. However, such a basis is unique for every combination of Ω and Ω'. Thus, we need

to use a global basis defined consistently for each direction Ω. For this, we choose the x-y basis, discussed above, and make the necessary coordinate transformations. Thus, we make the substitution

$$G(z,\Omega_k,\Omega_j) \rightarrow \ \mathcal{R}(\Omega_k,\Omega_j,\Omega_k)\mathbf{G}(\Omega_k,\Omega_j)\mathcal{R}^{-1}(\Omega_j,\Omega_k,\Omega_j) \tag{7}$$

where $\mathbf{G}(\Omega_j,\Omega_k)$ is the Mueller matrix differential scattering coefficient in the local coordinates (defined by the plane containing Ω_j and Ω_k) and $\mathcal{R}(\Omega_j,\Omega_k,\Omega')$ is a Mueller rotation matrix that rotates a Stokes vector from those local coordinates into the x-y coordinates defined for Ω'.

The second term on the right hand side of Eq. (2) describes the scattering away from the direction Ω. As such, the Mueller matrix for scattering only removes energy, based upon the Stokes vector of the incident radiance and should not otherwise retard, rotate, or depolarize the radiation in its original direction Ω. The matrix should thus appear as the differential matrix for diattenuation and only retain its polarization-dependent loss properties. The differential loss matrix can be defined from the scattering matrix[25] $\mathbf{m}$ by

$$\mathbf{\Lambda}(\mathbf{m}) \equiv \begin{pmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{01} & m_{00} & 0 & 0 \\ m_{02} & 0 & m_{00} & 0 \\ m_{03} & 0 & 0 & m_{00} \end{pmatrix}. \tag{8}$$

The following substitution should be made in the second term of the right hand side of Eq. (2):

$$G(z,\Omega_j,\Omega_k) \rightarrow \mathbf{\Lambda}[\mathcal{R}(\Omega_j,\Omega_k,\Omega_k)\mathbf{G}(z,\Omega_j,\Omega_k)\mathcal{R}^{-1}(\Omega_j,\Omega_k,\Omega_k)], \tag{9}$$

where we also include the rotation matrices as was done in Eq. (7).

Lastly, we need to consider the first term on right hand side of Eq. (2). The Mueller matrix absorption coefficient should also have the form of Eq. (8). For isotropic media, it is simply the product of the scalar absorption coeffient and the unit matrix. As for the phase function, we need to rotate the radiance into the local coordinate system of the scatterer, multiply by the absorption Mueller matrix, and then rotate back into the global coordinate system:

$$A(z,\Omega_j) \rightarrow \mathcal{R}(\Omega_j,\Omega_j,\Omega_j)\mathbf{A}(z,\Omega_j)\mathcal{R}^{-1}(\Omega_j,\Omega_j,\Omega_j), \tag{10}$$

where $\mathbf{A}(z,\Omega_j)$ is the Mueller matrix absorption function in the local coordinates. It should be noted that these rotations are not necessary for the phase functions implemented in this paper, since they do not depend upon incident direction.

When evaluating the BRDF at arbitrary incident-reflected direction combinations, we seek a suitable smoothing algorithm that is efficient and maintains the symmetries of the solution. The algorithm we chose is based upon interpolation by the Zernicke polynomials, similar to that used in Ref. 18.

3.3 Single scattering phase function

The scalar differential scattering coefficient $G(z,\Omega_j,\Omega_k)$ is the product of the single scattering phase function $p(z,\Omega',\Omega)$ and the scattering coefficient σ, where the phase function is normalized by

$$\int_{4\pi} p(z,\Omega',\Omega)\ \mathrm{d}\Omega' = 1. \tag{11}$$

We use a phase function based upon the scalar Henyey-Greenstein (HG) phase function,[26] given by

$$p(\alpha) = \frac{1}{4\pi}\frac{1-g^2}{(1-2g\cos\alpha+g^2)^{3/2}}, \tag{12}$$

where α is the scattering angle and the asymmetry parameter g corresponds to the average of the cosine of the scattering angle. For the polarimetric behavior, we assume that it has the same Henyey-Greenstein magnitude, but with a scattering matrix given in the local scattering coordinate system by

$$\mathbf{p}(\alpha) = p(\alpha) \times \begin{pmatrix} 1 & m_{01}(\alpha) & 0 & 0 \\ m_{01}(\alpha) & m_{11}(\alpha) & 0 & 0 \\ 0 & 0 & m_{22}(\alpha) & m_{23}(\alpha) \\ 0 & 0 & -m_{23}(\alpha) & m_{33}(\alpha) \end{pmatrix}. \tag{13}$$

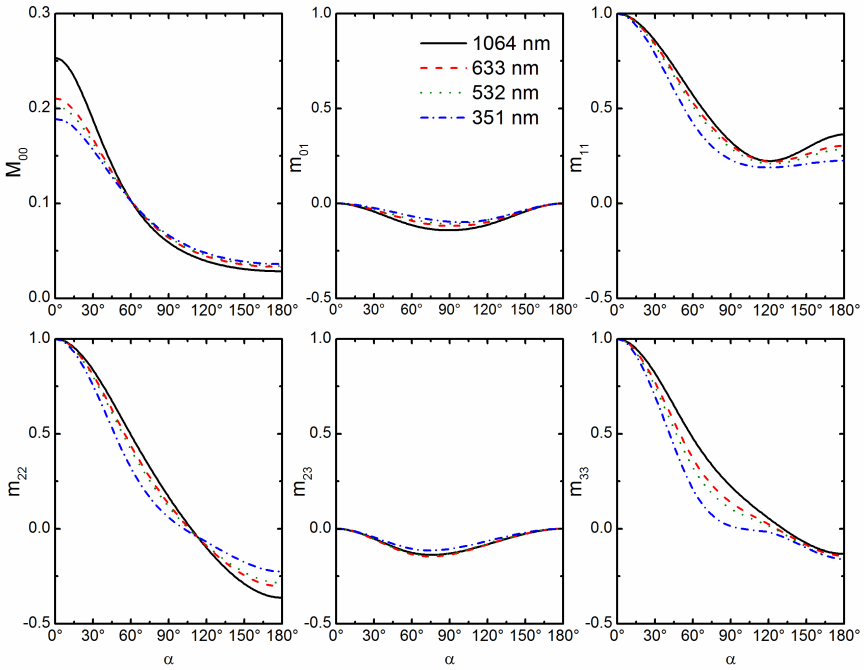


Figure 5. The polarimetric single scattering phase functions determined from the data.

This form is that expected for an isotropic medium. For each of the elements $m_{jk}(\alpha)$, we parameterize it with four points, $m_{jk}(0°)$, $m_{jk}(60\circ)$, $m_{jk}(120°)$, and $m_{jk}(180°)$, and use a cubic spline to interpolate it between these values, forcing $m'_{jk}(0°) = 0$ and $m'_{jk}(180°) = 0$. There are further restrictions on the form in the forward (0°) and backward (180°) scattering directions:

$$m_{11}(180°) = -m_{22}(180°), \tag{14a}$$

$$m_{01}(0°) = m_{01}(180°) = m_{23}(0°) = m_{23}(180°) = 0, \tag{14b}$$

$$m_{11}(0°) = m_{22}(0°). \tag{14c}$$

Finally, we found that we needed to fix the forward scattering matrix elements to $m_{11}(0°) = m_{22}(0°) = m_{33}(0°) = 1$ in order to obtain realizable matrices during data fitting.

4. RESULTS AND DISCUSSION

We performed a weighted Levenberg-Marquardt non-linear least squares fit of the data for each wavelength to the radiative transfer theory described above. The parameters included thirteen describing the phase function [g, $m_{11}(60°)$, $m_{11}(120°)$, $m_{11}(180°)$, $m_{22}(60°)$, $m_{22}(120°)$, $m_{01}(60°)$, $m_{01}(120°)$, $m_{23}(60°)$, $m_{23}(120°)$, $m_{33}(60°)$, $m_{33}(120°)$, and $m_{33}(180°)$], the index of refraction of the material, and the scattering coefficient. No absorption was included. Since the sample was thick, the results depended only weakly on the scattering coefficient; its inclusion in the fit was primarily to match the integrated reflectance. The weightings were chosen to be $(0.01 f_{\mathrm{r},00})^{-2}$ for the 00 terms and $(0.01)^{-2}$ for the normalized Mueller matrix elements. These fits were performed for several different choices of integration points (Ω_j and w_j) and found to be relatively stable. The resulting fits are shown in the second rows of Figs. 1–4. The root-mean-square (rms) differences between the data and the fits were 0.013 (351 nm), 0.008 (532 nm), 0.010 (633 nm), and 0.016 (1064 nm).

The average of the effective refractive index for all the fits was 1.018 with a standard deviation of 0.003. This value is very different than the index of refraction of solid PTFE (1.36). However, the main role of the refractive

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

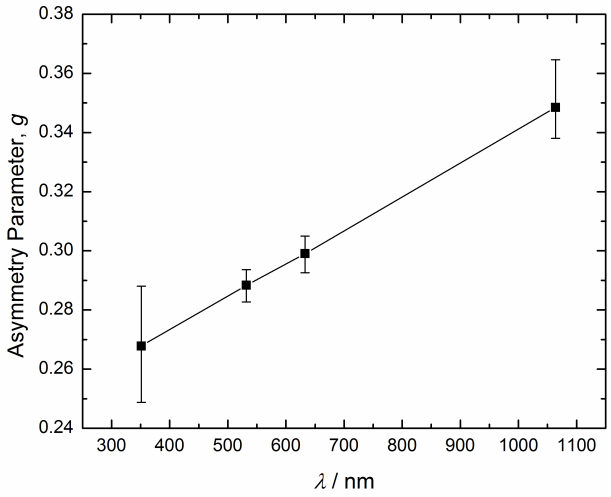


Figure 6. The asymmetry parameter g as a function of wavelength λ. The points represent the mean of five different fits, each using a different order of Gaussian cubature for the sampled integration points, and the uncertainties show the entire range of the results.

index in these calculations is the application of Snell's law in relating external and internal propagation angles. Large amounts of roughness substantially decreases the average angle of refraction.[27] Thus, this index should not be thought of as the index of PTFE, but rather an indication of how rough the material is.

The resulting phase functions are shown in Fig. 5, and the parameter g is shown as a function of wavelength in Fig. 6. The values shown in Fig. 6 represent the mean of the values obtained from the different integration points, while the uncertainties shown represent the entire range of the values. There are some noticeable trends in the results. As the wavelength increases, the single scattering phase function becomes more strongly peaked in the forward scattering direction (i.e., g increases with wavelength λ). As the wavelength increases, the amount of depolarization also decreases. That is, for most scattering angles, $|m_{jk}|$ increases with wavelength.

The values obtained for the asymmetry parameter g (see Fig. 6) are significantly smaller than those observed in other studies. For example, Huber *et al.* found that, at 633 nm, $g = 0.89 \pm 0.03$, and at 350 nm, $g = 0.710 \pm 0.004$.[28] We believe that this may be partly due to material-to-material variations and partly due to the differences in analysis. Huber *et al.* used much thinner films, laser-ablated them to thickness, and appear to use the natural index of refraction of the PTFE (1.36) in their analysis, thus possibly ignoring the effects of roughness and overestimating their value of g. But, by measuring the scattering only in reflection, albeit for large incident angles, our method may be unduly weighting the results by the shape of the phase function in the backwards scattering direction. That is, the scattering can be considered as a sum of singly scattered and multiply scattered radiation. The former adds structure to the BRDF and the Mueller matrix, while the later is primarily Lambertian and depolarizing. Insofar as we are using the Henyey-Greenstein form for the phase function, we are restricting ourselves to fitting the behavior at large angles, rather than at its peak in the forward scattering direction.

REFERENCES

[1] Weidner, V. R. and Hsia, J. J., "Reflection properties of pressed polytetrafluoroethylene powder," J. Opt. Soc. Am. **71**, 856–861 (1981).

[2] Weidner, V. R., Hsia, J. J., and Adams, B., "Laboratory intercomparison study of pressed polytetrafluoroethylene powder," Appl. Opt. **24**, 2225–2230 (1985).

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

[3] Barnes, P. Y., and Hsia, J. J., "45°/0° Reflectance Factors of Pressed Polytetrafluoroethylene (PTFE) Powder," NIST Technical Note 1413 (NIST, Gaithersburg, 1995).

[4] Betty, C.L., Fung, A.K., and Irons, J., "The measured polarized bidirectional reflectance distribution function of a Spectralon calibration target," *Geoscience and Remote Sensing Symposium, 1996. IGARSS '96. 'Remote Sensing for a Sustainable Future.'*, vol.4, pp. 2183–2185 (1996).

[5] Haner, D.A., McGuckin, B.T., and Bruegge, C.J., "Polarization characteristics of Spectralon illuminated by coherent light," Appl. Opt. **38**, 6350–6356 (1999).

[6] Goldstein, D. H., Chenault, D. B., and Pezzaniti, L., "Polarimetric characterization of Spectralon," in *Polarization: Measurement, Analysis, and Remote Sensing II*, Proc. SPIE **3754**, 126–136 (1999).

[7] Georgiev, G.T., and Butler, J.J., "The effect of incident light polarization on spectralon BRDF measurements," in *Sensors, Systems, and Next-Generation Satellites VIII*, in *Sensors, Systems, and Next-Generation Satellites VIII*, Proc. SPIE **5570**, 492–502 (2004)

[8] Tsai, B. K., Allen, D. W., Hanssen, L. M., Wilthan, B., and Zeng, J., "A comparison of optical properties between solid PTFE (Teflon) and (low density) sintered PTFE," in *Reflection, Scattering, and Diffraction from Surfaces*, Proc. SPIE **7065**, 70650Y (2008).

[9] Bhandari, A., Hamre, B., Frette, Ø., Zhao, L., Stamnes, J. J., and Kildemo, M., "Bidirectional reflectance distribution function of Spectralon white reflectance standard illuminated by incoherent unpolarized and plane-polarized light," Appl. Opt. **50**, 2431–2442 (2011).

[10] Bruegge, C. J., Stiegman, A. E., Rainen, R. A., and Springsteen, A. W., "Use of Spectralon as a diffuse reflectance standard for in-flight calibration of earth-orbiting sensors," Opt. Eng. **32**, 805–814 (1993).

[11] Goldstein, D.H., and Chenault, D.B., "Spectropolarimetric reflectometer," Opt. Eng. **41**, 1013–1020 (2002).

[12] Noble, H., Lam, W.-S. T., Smith, G., McClain, S., and Chipman, R.A., "Polarization scattering from a Spectralon calibration sample," in *Polarization Science and Remote Sensing III*, Shaw, J.A., and Tyo, J.S., eds., Proc. SPIE **6682**, 668219 (2007).

[13] Germer, T. A., and Patrick, H. J., "Mueller matrix bidirectional reflectance distribution function measurements and modeling of diffuse reflectance standards," in *Polarization Science and Remote Sensing V*, Proc. SPIE **8160**, 81600D (2011).

[14] Svensen, Ø., Kildemo, M., Maria, J., Stamnes, J.J., and Frette, Ø., "Mueller matrix measurements and modeling pertaining to Spectralon white reflectance standards," Opt. Exp. **20**, 15045–15053 (2012).

[15] Sanz, J. M., Extremiana, C., and Saiz, J. M., "Comprehensive polarimetric analysis of Spectralon white reflectance standard in a wide visible range," Appl. Opt. **52**, 6051–6062 (2013).

[16] Höpe, A., and Hauer, K.-O., "Three-dimensional appearance characterization of diffuse standard reflection materials," Metrologia **47**, 295–304 (2010).

[17] Patrick, H. J., Hanssen, L., Zeng, J., and Germer, T. A., "BRDF measurements of graphite used in high-temperature fixed point blackbody radiators: a multi-angle study at 405 nm and 658 nm," Metrologia **49**, S81–S92 (2012).

[18] Germer, T. A., "Full four-dimensional and reciprocal Mueller matrix bidirectional reflectance distribution function of sintered polytetrafluoroethylene," Appl. Opt. **56**, 9333–9340 (2017).

[19] Koenderink, J. J., and van Doorn, A. J., "Phenomenological description of bidirectional surface reflection," J. Opt. Soc. Am. A. **15**, 2903–2912 (1998).

[20] Germer, T. A., and Asmail, C. C., "Goniometric optical scatter instrument for out-of-plane ellipsometry measurements," Rev. Sci. Instrum. **70**, 3688–3695 (1999).

[21] Flynn, D. S., and Alexander, C., "Polarized surface scattering expressed in terms of a bidirectional reflectance distribution function matrix," Opt. Eng. **34**, 1646–1650 (1995).

[22] Compain, E., Poirier, S., and Drevillon, B., "General and Self-Consistent Method for the Calibration of Polarization Modulators, Polarimeters, and Mueller-Matrix Ellipsometers," Appl. Opt. **38**, 3490–3502 (1999).

[23] Germer, T. A., and Patrick, H. J., "Effect of bandwidth and numerical aperture in optical scatterometry," in *Metrology, Inspection, and Process Control for Microlithography XXIV*,Proc. SPIE **7638**, 76381F-1–10 (2010).

[24] Hapke, B., *Theory of Reflectance and Emittance Spectroscopy*, pp. 175–176 (Cambridge University, Cambridge, 1993).

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

[25] Azzam, R. M. A., "Propagation of partially polarized light through anisotropic media with and without depolarization: A differential 4×4 matrix calculus," J. Opt. Soc. Am. **68**, 1756–1767 (1978).
[26] Henyey, L. G. and Greenstein, J. L., "Diffuse radiation in the Galaxy," Astrophys. J. **93**, 70–83 (1941).
[27] Germer, T. A., "Polarized light diffusely scattered under smooth and rough surfaces," in *Polarization Science and Remote Sensing*, J. A. Shaw and J. S. Tyo, Eds., Proc. SPIE **5158**, 193-204 (2003).
[28] Huber, N., Heitz, J., and Bäuerle, D., "Pulsed-laser ablation of polytetrafluoroethylene (PTFE) at various wavelengths," Eur. Phys. J. Appl. Phys. **25**, 33–38 (2004).

Germer, Thomas.
"Polarized single-scattering phase function determined for a common reflectance standard from bidirectional reflectance measurements."
Paper presented at SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, FL, United States. April 15, 2018 - April 19, 2018.

The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST

Speaker/Author: Michael A. Lombardi
Time and Frequency Division
National Institute of Standards and Technology (NIST)
Boulder, Colorado, United States
lombardi@nist.gov

Abstract: The National Institute of Standards and Technology (NIST) has provided remote frequency and time calibration services to customers for more than three decades. These services continuously compare a customer's primary frequency and/or time standard to the coordinated universal time scale kept at NIST, known as UTC(NIST), which is the U. S. national standard for frequency and time. The remote calibration services differ from traditional calibration services in at least two important ways. The first difference is that the customer does not send the device under test to NIST. Instead, NIST sends equipment to the customer that automates the measurements and returns the results via a network connection. The second difference is that the calibration never stops. New measurement results are recorded 24 hours per day, 7 days a week. This allows customers to continuously establish traceability to the International System (SI) via UTC(NIST) without ever disturbing or moving their standard.

Since their inception, these services have evolved to meet the frequency and time requirements of customers in a variety of public and private sectors. Initially, most of the customers were calibration and metrology laboratories, primarily located at U. S. military installations and at defense contractor sites. Those customers remain an integral part of the customer base, but the reach of the services now extends to research laboratories, the aerospace industry, the energy industry, electronics and instrument manufacturers, and most recently, to financial markets and stock exchanges. This paper discusses the reach and impact of the NIST remote frequency and time calibration services by describing how they work, their calibration and measurement capabilities, their quality system, and the requirements of the customers that they serve.

1. Introduction and Overview

Although they are now classified as calibration services, the services that comprise the remote frequency and time calibration program at NIST have their roots in the measurement assurance program (MAP) that was formally announced by NIST's predecessor, the National Bureau of Standards (NBS), in the early 1980s. A MAP was defined as a "quality assurance program for a measurement process that quantifies the total uncertainty of the measurements with respect to national or other designated standards and demonstrates that the total uncertainty is small enough to meet the user's requirements." A MAP differed from a calibration service because it focused on "the quality of the measurements being made in the participating laboratory rather than on the properties of participants instruments or standards." Participation in a MAP was potentially a way to calibrate "the entire laboratory" [1].

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

Before implementing the MAP program, the NBS time and frequency division had little experience performing remote calibrations for customers. The division did have, however, many years of experience in frequency and time research, in distributing frequency and time reference signals, and in participating in national and international comparisons. Once the division was aware of the requirements of potential customers, it combined the experience gained from past research, distribution, and comparison efforts to develop new services. In fact, each of the remote frequency and time calibration services currently offered by NIST can be traced to technology that was originally developed for other projects or experiments, and that was then refined until it was reliable and "user friendly" enough to make it available to customers. The goal was to make the service accessible to any customer with a frequency or time measurement requirement, regardless of where they were located or their level of metrology experience.

To illustrate this, note that the Frequency Measurement and Analysis Service (FMAS), which debuted in 1984, has its origins in work performed to develop frequency measurement systems for the United States Air Force from about 1980 to 1983 [2]. The Time Measurement and Analysis Service (TMAS), which debuted in 2006 [3], is based on technology first developed at NIST in 2005 to support real-time, international time comparisons for SIM (Sistema Interamericano de Metrologia or Interamerican Metrology System in English), the regional metrology organization (RMO) that supports North, Central, and South America [4]. The common-view disciplined clock services [5] that first appeared as optional add-ons to the TMAS in 2010, were the outgrowth of experiments conducted with the United States Coast Guard beginning in 2008 [6]. Finally, the time code output and time code monitoring options for the TMAS have their roots in development work first done at NIST to support international comparisons of Internet time servers in the SIM region in 2014 [7].

NIST's approach to a MAP for remote frequency and time calibrations has always been based on a consistent philosophy and mode of operation. Every customer receives an automated measurement system that runs continuously with very little attention. This has several benefits. First, automation and standardized equipment means that every customer makes their measurements the exact same way. Second, the measurements never stop, continuing 24 hours per day, seven days per week. Third, the customer's standard is never sent outside their laboratory, making it always available for use. Fourth, because the measurements are automated, labor costs are reduced. Finally, and perhaps most importantly, NIST staff monitor and fully support the measurements. They are available to answer questions and provide technical support, to help customers best utilize the measurement system within the framework of their laboratory, and to educate customers about time and frequency metrology. This results in services with more "value added" than traditional calibration services, which typically just deliver a certificate or report.

Table 1 lists each remote frequency and time calibration service and summarizes its measurement capabilities. Section 2 provides a technical description of how the services work. The customers and industries reached by the services and their metrological requirements are discussed in Section 3. The NIST quality system that supports the services is discussed in Section 4. Customer accreditation is discussed in Section 5 and Section 6 provides a summary.

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

Table 1. The NIST remote frequency and time calibration services.

Service Name	Frequency Measurement and Analysis Service (FMAS)	Time Measurement and Analysis Service (TMAS)	NIST Disciplined Clock Service (NISTDC)		Time Code Output Service	Time Code Monitoring Service
NIST ID Number	76100C	76101C	76102C, Rubidium	76103C, Cesium	76104C	76105C
Year Introduced	1984	2006	2010		2017	2017
Other Services Required	None	None	76101C, customer must supply cesium clock for 76103C		76101C, either 76102C or 76103C	76101C, either 76102C or 76103C
Signals Measured	All frequencies from 1 Hz to 120 MHz	1 Hz	1 Hz		NA	NTP packets
Measurement Channels	5	1	1		2 outputs, NTP/PTP	12
Time Uncertainty ($k = 2$)	NA	12 ns (typical)	10 ns (typical)		NA	< 10 µs, but usually larger due to network asymmetry
Frequency Uncertainty ($k = 2$)	2×10^{-13} at 1 day	5×10^{-14} at 1 day	5×10^{-15} at 1 day		NA	NA
Interval between comparison graphs	24 hours	10 minutes	10 minutes		NA	10 minutes
Calibration Report Interval	Monthly	Monthly	Monthly		NA	NA, results available via Internet

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

2. Technical Description

The foundation for the remote calibration services is built upon a few fundamental time and frequency metrology techniques; including time interval measurements, common-view observations of Global Positioning System (GPS) satellite signals, disciplining a local clock so that its frequency and time agrees with a remote reference, and comparing received time codes to a reference clock. This section briefly describes each technique and how it applies to the services.

The FMAS (76100C) and TMAS (76101C) measurement systems include a time interval counter (TIC) and implement the time interval measurement technique. Figure 1 shows a simplified configuration where a device under test (DUT) is compared to a local reference. In both the FMAS and TMAS systems, the DUT signal is a 1 Hz output from the customer's primary standard and the local reference signal is the 1 Hz output of a GPS timing receiver. Figure 1 shows frequency dividers in the path between the signal sources and the TIC, but in the case of the TMAS no frequency division is necessary. The FMAS does, however, require frequency division, a feature that allows it to measure a large range of frequencies but precludes it from measuring the DUT's time offset from NIST. The FMAS divides each signal to 1 Hz before it reaches the TIC, and includes a multiplexer so that up to five signals can be read in sequence. Because time interval is the reciprocal of frequency, the time interval collected by either the FMAS or TMAS can easily be converted to frequency data [2].

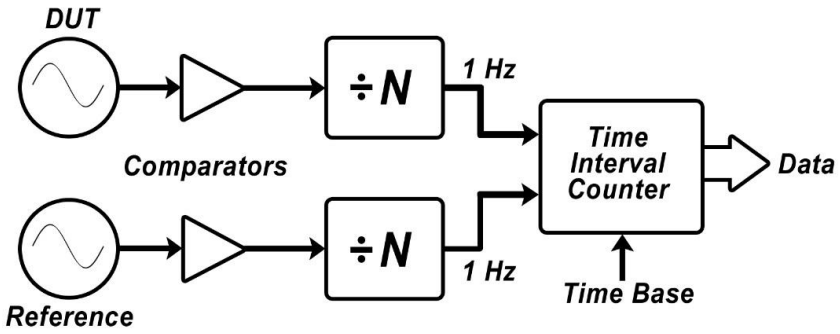


Figure 1. Time Interval Measurement System with Frequency Dividers.

The common-view GPS measurement technique is an essential part of the remote calibration services. The technique was first demonstrated at NBS in 1980 [8] and remains the primary method used by national metrology institutes for sending data to the BIPM (Bureau International des Poids et Mesures (BIPM) in French or International Bureau of Weights and Measures in English) for the calculation of Coordinated Universal Time (UTC) [9]. Two clocks located at different locations cannot be directly compared, but common-view allows them to be indirectly compared if a common signal that serves as a transfer standard can be received at both clock sites. The GPS signals provide this transfer standard.

Figure 2 is a diagram of a common-view comparison between the customer's standard and the UTC(NIST) time scale. The customer and NIST simultaneously compare their standards to signals from GPS satellites and then exchange the measurement results via the Internet, so they

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

can be subtracted from each other. Computing the difference between the two measurements removes the effects of GPS and produces an estimate of the difference between the customer's standard and UTC(NIST).

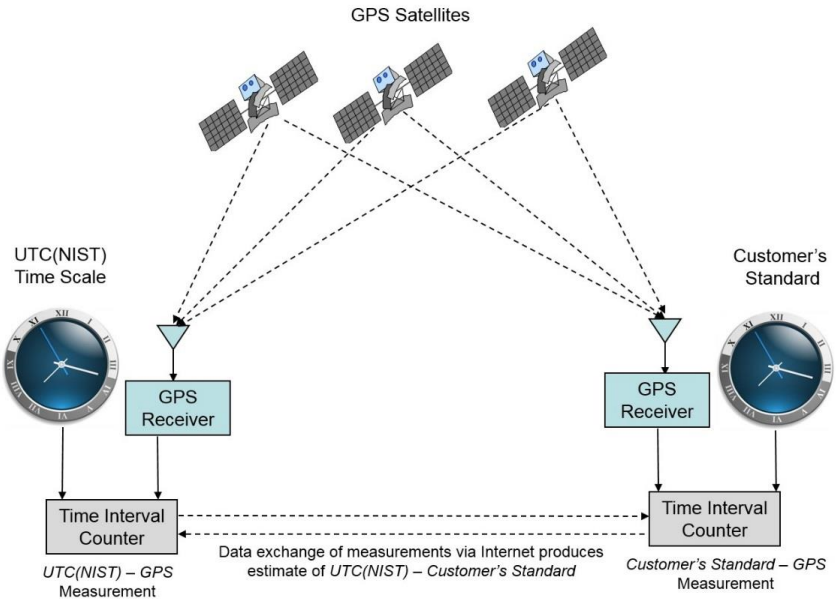


Figure 2. A common-view GPS comparison between a customer and NIST.

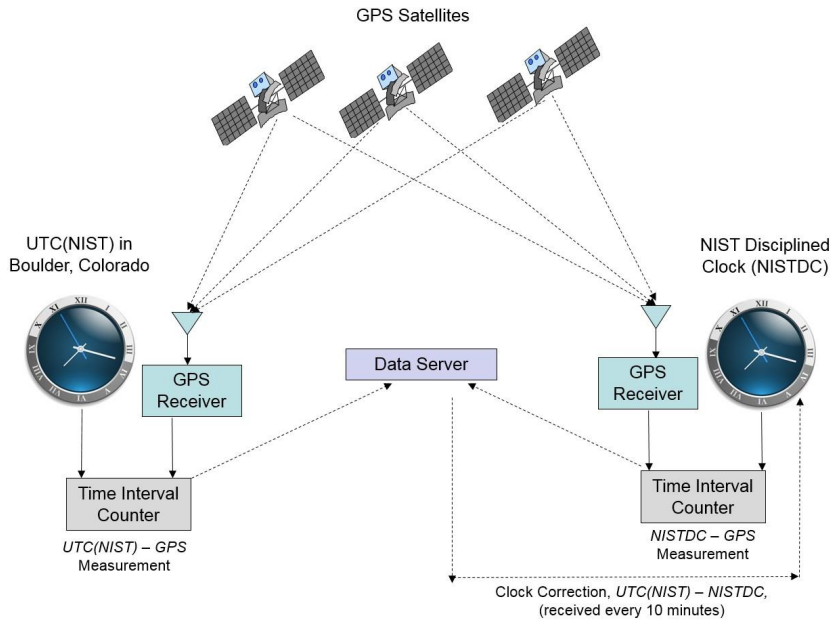


Figure 3. A common-view GPS comparison resulting in a NIST disciplined clock.

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

The TMAS produces the *UTC(NIST) - Customer's Standard* time difference estimate every 10 minutes (Figure 2), and the NIST disciplined clock (NISTDC) service converts these measurements to frequency corrections that are then applied to the customer's standard to keep it locked to UTC(NIST), as shown in Figure 3. The standard can be either a rubidium clock supplied by NIST (76102C) or a cesium clock supplied by the customer (76103C).

Some NIST customers, especially those in the financial sector, require the highly accurate time they receive from NIST to be distributed to computer systems within their organization or network, so that transactions and files can be accurately time stamped. These customers use the NISTDC, which typically keeps time within 10 ns of the UTC(NIST) time scale, as the reference for the Time Code Output service (76104C), which can distribute time codes to the customer's server or client computers via either the Network Time Protocol (NTP) or Precision Time Protocol (PTP).

Another requirement of financial sector customers is checking the accuracy of their time servers, so that they can ensure their time codes were correct not only when transmitted, but also when received. The Time Code Monitoring service (76105C) can check the accuracy from as many as 12 of the customer's servers by requesting NTP packets and comparing them to the time kept by the NISTDC. The uncertainty of this measurement is limited by network asymmetry, but this problem is largely eliminated by locating the NISTDC as close as possible to the servers being monitored, typically placing it on the same local area network in the same data center.

The TMAS and each of its optional services can be delivered to customers who install a single rack-mounted instrument (Figure 4). The instrument includes the GPS receiver and time interval counter required by the TMAS, the rubidium oscillator required by the NISTDC, the NTP/PTP server required by the Time Code Output service, and a hardware clock required by the Time Code Monitoring service.

Figure 4. The TMAS measurement system that NIST supplies to customers.

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

3. Customer Locations, Industries Reached, and Customer Requirements

Because GPS signals are globally available, a common-view link to NIST can be established anywhere on Earth. At this writing (March 2018), NIST provides services to more than 50 remote calibration sites, as shown on the map in Figure 5. The red clocks on the map represent FMAS customers, the blue clocks represent TMAS customers, and the green clocks represent TMAS/NISTDC customers. The customers are located in 23 states and seven sites outside of the U. S.

Figure 5. Locations of NIST remote frequency and time calibration customers (red clocks indicate FMAS, blue clocks indicate TMAS, green clocks indicate TMAS/NISTDC).

The services reach and impact a wide variety of industrial sectors (Figure 6). The largest customer group consists of in-house calibration laboratories who use the NIST reference to calibrate their primary standard and then use their primary standard as the reference for calibrations of other equipment. Industries whose in-house calibration laboratories rely on the NIST services include U. S. defense contractors (14%), aerospace (10%), and nuclear energy (2%), in addition to U. S. military installations (14%). Collectively, these in-house calibration laboratories represent 40% of the customer base. The largest industrial sector represented in the customer base consists of manufacturers of electronic test and measurement equipment and instrumentation (28%), including frequency and time standards, such as atomic oscillators and GPS disciplined oscillators. These customers require NIST validation of their measurements to ensure that the products they design, manufacture, and sell can meet their desired specifications for frequency and/or time. Private calibration laboratories that sell calibrations to customers outside of their own organizations, and non-military U. S. government laboratories each contribute 8% to the NIST customer total. The customers described in this paragraph are primarily interested in frequency measurements, although many also maintain an accurate time standard.

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

The financial sector currently represents 16% of the customer base, a percentage that is likely to increase in the future. Financial market customers generally have little interest in frequency measurements. Their main concern is obtaining accurate time from UTC(NIST) for the time stamping of transactions. The importance of this sector in economic terms is difficult to overstate, and thus continuing to meet the needs of these customers is a priority for the remote calibration program.

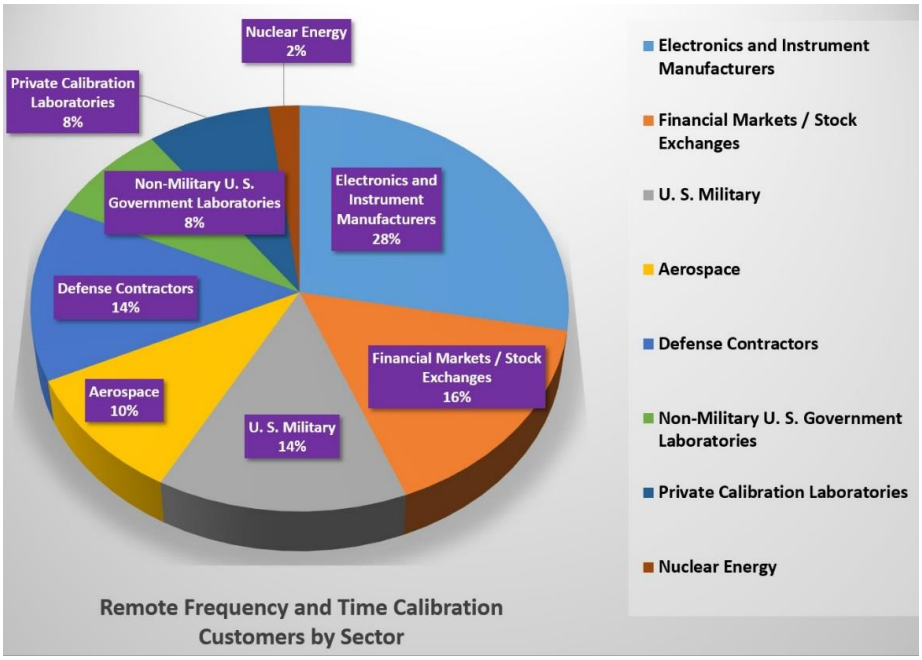


Figure 6. Percentages of NIST remote frequency and time calibration customers by sector.

The customers consider the measurement assurance they receive from NIST to be an essential part of their daily operation. In other words, they need the capabilities of the NIST services, and NIST must ensure that these capabilities meet or exceed the customer's requirements. These requirements vary from sector to sector. For example, a frequency measurement uncertainty requirement of a few parts in 10^{13} over a 1-day interval is currently small enough to meet the needs of most equipment manufacturers and both in-house and private calibration laboratories. This requirement is met by both the FMAS and TMAS (Table 1). Industrial requirements for time measurement uncertainty are generally limited to ~1 µs, with the most stringent customer requirements currently about 0.1 µs (100 ns), which is also easily exceeded by the TMAS and NISTDC services. Financial market timing requirements have uncertainty requirements that are typically 1000× larger, but that can be difficult to meet because they apply not only to the electrical pulses used for synchronization, but also to the digital time codes that must be transferred to computer systems. The most stringent financial requirement in place at this writing (March 2018) is 100 µs time accuracy (with respect to UTC) for clocks involved in high frequency trading in Europe [10]. This requirement is currently being met and exceeded by the TMAS and its optional services [11]. Although customer requirements are currently being met,

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

enhancing the services to meet future requirements has been, and remains, a continuous challenge.

4. NIST Quality System

The remote frequency and time calibration services are supported by the NIST quality system (NIST QS). The calibration part of the NIST QS is built on the foundation of *ISO Guide 17025* [12]. Measurements uncertainty is reported to customers using methods compliant with the international standard [13], but the uncertainty analysis also incorporates metrics specific to frequency and time metrology, such as the Allan deviation and Time deviation [14].

The NIST QS documentation includes the NIST quality manual (QM-I), a quality manual for each of the technical divisions at NIST (QM-IIs), and a quality manual for each of the calibration services (QM-IIIs) [15, 16]. The time and frequency division maintains a QM-II quality manual, and separate QM-III quality manuals for the FMAS and TMAS services. These manuals are revised whenever necessary. In addition, quality reports that document any changes or issues related to the services are required to be written every three months. These reports are sent to the NIST quality manager and distributed to other NIST managers as necessary.

Each technical division at NIST has its quality system assessed every five years. The time and frequency division was previously assessed in 2005, 2010, and 2015, with the next assessment scheduled for 2020. After all findings recorded by the assessors are addressed (findings can include either non-conformities or observations), the NIST quality manager presents the division's quality system at a SIM general assembly meeting, where it can be approved or disapproved. If approval is given, it is valid for five more years.

NIST and other national metrology institutes with quality systems approved by their RMO can apply to have their calibration and measurement capabilities (CMCs), and the services that provide them, listed in the BIPM's Key Comparison Database (KCDB) [17, 18]. Inclusion in the KCDB means that the calibration service is internationally recognized by all signatories of the CIPM (Comité International des Poids et Mesures in French or International Committee of Weights and Measures in English) mutual recognition agreement (MRA). The MRA signatories include nearly every industrial nation [19]. The FMAS and TMAS are each listed in the KCDB. This means that their measurements are internationally recognized as being traceable to the SI at their stated levels of uncertainty.

5. Customer Accreditation

Most NIST customers must have their measurement capabilities periodically assessed or audited, which is often a prerequisite for them staying in business or meeting the contractual requirements of their own customers. Many NIST customers also seek laboratory accreditation. Having access to the continuous measurements records that both the FMAS and TMAS provide benefits the customer and can make the assessment and accreditation processes go much smoother. For example, the two major accreditation bodies for U. S. calibrations laboratories are the National Voluntary Laboratory Accreditation Program (NVLAP) and the American Association for Laboratory Accreditation (A2LA). As of March 2018, five NIST customers are accredited for

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

frequency measurements by NVLAP [20] and nine are accredited by A2LA [21] with CMCs ranging from 2.5×10^{-12} to 1.3×10^{-13}.

Customers outside of the traditional calibration world, such as those in the financial sector, also undergo periodic assessments and audits and must be able to prove that they comply with the standards of regulatory bodies. In the financial sector, these regulatory bodies include the Financial Industry Regulatory Authority (FINRA) and the ultimate authority, the U. S. Securities and Exchange Commission (SEC). The TMAS/NISTDC service, which can provide logs of time measurements results recorded every second, allows financial market customers to easily provide regulators with the metrological evidence they are seeking.

6. Summary

By providing continuous monitoring and reporting of a customer's measurements, the NIST remote frequency and time calibration services do more than traditional calibration services; they provide measurement assurance and improve the quality of all frequency and time measurements made at the customer's location. The services satisfy the requirements of customers in many sectors and industries, are supported by the NIST quality system, are internationally recognized as being traceable to the SI, and can help customers pass legal metrology assessments and achieve accreditation.

7. References

[1] B. Belanger, "Measurement Assurance Programs – Part I: General Introduction," *NBS Special Publication 676-I*, 65 p., May 1984.

[2] M. A. Lombardi, "Remote Frequency Calibrations: The NIST Frequency Measurement and Analysis Service," *NIST Special Publication 250-29*, 90 p., June 2004.

[3] M. A. Lombardi and A. N. Novick, "Remote Time Calibrations via the NIST Time Measurement and Analysis Service," *NCSLI Measure J. Meas. Sci.*, vol. 1, no. 4, pp. 50-59, December 2006.

[4] M. Lombardi, A. Novick, J. M. López-Romero, J. S. Boulanger, and R. Pelletier, "The inter-American metrology system (SIM) common-view GPS comparison network," *Proceedings of the 2005 IEEE International Frequency Control Symposium*, pp. 691–698, August 2005.

[5] M. A. Lombardi, "A NIST disciplined oscillator: delivering UTC(NIST) to the calibration laboratory," *NCSLI Measure J. Meas. Sci.*, vol. 5, no. 4, pp. 56-64, December 2010.

[6] M. A. Lombardi and A. P. Dahlen, "A common-view disciplined oscillator," *Rev. Sci. Instrum.*, vol. 81, no. 5, pp. 055110-1/6, May 2010.

[7] M. Lombardi, J. Levine, J. Lopez, F. Jimenez, J. Bernard, M. Gertsvolf, H. Sanchez, O. G. Fallas, L. C. Hernández Forero, R. José de Carvalho, M. N. Fittipaldi, R. F. Solis, and F. Espejo, "International Comparisons of Network Time Protocol Servers," *Proceedings of the*

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

Precise Time and Time Interval Meeting (ION PTTI), Boston, Massachusetts, USA, pp. 57-66, December 2014.

[8] D. W. Allan and M. A. Weiss, "Accurate Time and Frequency Transfer During Common-View of a GPS Satellite," *Proceedings of the 1980 Frequency Control Symposium*, pp. 334-346, May 1980.

[9] Bureau International des Poids et Mesures (BIPM), *BIPM Annual Report on Time Activities*, vol. 11, 134 p., 2016.

[10] "Commission Delegated Regulation (EU) 2017/574 of 7 June 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council with regard to regulatory technical standards for the level of accuracy of business clocks," 2017, *Official Journal of the European Union*, L 87/148-151, March 2017.

[11] M. Lombardi, A. Novick, B. Cooke, and G. Neville-Neil, "Accurate, Traceable, and Verifiable Time Synchronization for World Financial Markets," *Journal of Research of the National Institute of Standards and Technology*, vol. 121, pp. 436-463, 2016.

[12] International Organization for Standardization (ISO), "General Requirements for the Competence of Testing and Calibration Laboratories," *ISO/IEC Guide 17025*, 2005.

[13] JCGM, "Evaluation of measurement data—Guide to the expression of uncertainty in measurement," *JCGM 100*, 2008.

[14] IEEE, "Standard definitions of physical quantities for fundamental frequency and time metrology—Random instabilities," *IEEE Standard 1139*, 2008.

[15] NIST Quality System web portal (https://www.nist.gov/nist-quality-system).

[16] S. Bruce, "The NIST Quality System for Measurement Services: A Look at its Past Decade and a Gaze towards its Future," *Proceedings of the 2013 NCSL International Workshop and Symposium*, July 2013.

[17] C. Thomas and A. J. Wallard, "A User's Guide to the Information in the BIPM Key Comparison Database," *NCSLI Measure J. Meas. Sci.*, vol. 2, no. 4, pp. 22-27, December 2007.

[18] BIPM Key Comparison Database (https://kcdb.bipm.org/AppendixC/).

[19] CIPM Mutual Recognition Agreement Database (https://www.bipm.org/en/cipm-mra/).

[20] NVLAP Interactive Web System (NIWS) and Directory of Accredited Laboratories (https://www-s.nist.gov/niws/index.cfm).

[21] A2LA Directory of Accredited Organizations (https://portal.a2la.org/search/).

Lombardi, Michael.
"The Reach and Impact of the Remote Frequency and Time Calibration Program at NIST."
Paper presented at 2018 NCSL International Workshop and Symposium, Portland, OR, United States. August 27, 2018 - August 30, 2018.

Three-Party Quantum Self-Testing with a Proof by Diagrams

Spencer Breiner
NIST
spencer.breiner@nist.gov

Amir Kalev
Joint Center for Quantum Information and Computer Science (QuICS)
amirk@umd.edu

Carl A. Miller
NIST
Joint Center for Quantum Information and Computer Science (QuICS)
camiller@umd.edu

Quantum self-testing addresses the following question: is it possible to verify the existence of a multipartite state even when one's measurement devices are completely untrusted? This problem has seen abundant activity in the last few years, particularly with the advent of parallel self-testing (i.e., testing several copies of a state at once), which has applications not only to quantum cryptography but also quantum computing. In this work we give the first error-tolerant parallel self-test in a three-party (rather than two-party) scenario, by showing that an arbitrary number of copies of the GHZ state can be self-tested. In order to handle the additional complexity of a three-party setting, we use a diagrammatic proof based on categorical quantum mechanics, rather than a typical symbolic proof. The diagrammatic approach allows for manipulations of the complicated tensor networks that arise in the proof, and gives a demonstration of the importance of picture-languages in quantum information.

1 Introduction

Quantum rigidity has its origins in quantum key distribution, which is one of the original problems in quantum cryptography. In the 1980's Bennett and Brassard proposed a protocol for secret key distribution across an untrusted public quantum channel [3]. A version of the Bennett-Brassard protocol can be expressed as follows: Alice prepares N EPR pairs, and shares the second half of these pairs with Bob through the untrusted public channel. Alice and Bob then perform random measurements on the resulting state, and check the result to verify that indeed their shared state approximates N EPR pairs. If these tests succeed, Alice and Bob then use other coordinated measurement results as the basis for their shared key. Underlying the proof of security for the Bennett-Brassard protocol is the idea that if a shared 2-qubit state approximates the behavior of a Bell state under certain measurements, then the state itself must itself approximate a Bell state.

If we wish to deepen the security, we can ask: what if Alice's and Bob's measurement devices are also not trusted? Can we prove security at a level that guards against possible exploitation of defects in their measurement devices? This leads the question of quantum rigidity: is it possible to completely verify the behavior of n untrusted quantum measurement devices, based only on statistical observation of their measurement outputs, and without any prior knowledge of the state they contain?

We say that a n-player cooperative game is *rigid* if an optimal score at that game guarantees that the players must have used a particular state and particular measurements. We say that a state is *self-testing* if its existence can be guaranteed by such a game. Early results on this topic focused on self-testing the 2-qubit Bell state [31, 19, 24]. Since then a plethora of results on other games and other states have appeared. The majority of works have focused on the bipartite case, and there are a smaller number of works that address n-partite states for $n \geq 3$ [25, 21, 38, 37, 29, 14].

More recently, it has been observed that rigid games exist that self-test not only one copy of a bipartite state, but several copies at once. Such games are a resource not only for cryptography, but also for

Submitted to:
QPL 2018

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

quantum computation: these games can be manipulated to force untrusted devices to perform measurements on copies of the Bell state which carry out complex circuits. This idea originated in [32] and has seen variants and improvements since then [21, 26, 11]. For such applications, it is important that the result include an error term which is (at most) bounded by some polynomial function of the number of copies of the state.[1]

It is noteworthy that all results proved so far for error-tolerant self-testing of several copies of a state at once (that is, parallel self-testing) apply to bipartite states only [22, 23, 13, 26, 27, 28, 5, 32, 12, 8, 10, 9]. There is a general multipartite self-testing result in [36] which can be applied to the parallel case, but it is not error-tolerant and no explicit game is given. Complexity of proofs is a factor in establishing new results in this direction: while it would be natural to extrapolate existing parallelization techniques to prove self-tests for n-partite states, the proofs for the bipartite case are already difficult, and we can expect that the same proofs for $n \geq 3$ are more so. Yet, if this is an obstacle it is one worth overcoming, since multi-partite states are an important resource in cryptography. For example, a much-cited paper in 1999 [15] proved that secret sharing is possible using several copies of the GHZ state $|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, in analogy to the use of Bell states in the original QKD protocol [3].

In this work, we give the first proof of an error-tolerant parallel *tripartite* self-test. Specifically, we prove that a certain class of 3-player games self-test N copies of the GHZ state, for any $N \geq 1$, with an error term that grows polynomially with N. To accomplish this we introduce, for the first time, the graphical language of categorical quantum mechanics into the topic of rigidity. As we will discuss below, the use of graphical languages is a critical feature of the proof — games involving more than 2 players involve complicated tensor networks, which are not easily expressed symbolically. Our result thus demonstrates the power and importance of visual formal reasoning in quantum information processing.

1.1 Categorical quantum mechanics

Category theory is a branch of abstract mathematics which studies systems of interacting processes. In Categorical Quantum Mechanics (CQM), categories (specifically *symmetric monoidal categories*) are used to represent and analyze the interaction of quantum states and processes.

Inspired by methods from computer science, CQM introduces a explicit distinction between traditional quantum semantics in Hilbert spaces and the syntax of quantum protocols and algorithms. In particular, symmetric monoidal categories support a diagrammatic formal syntax called *string diagrams*, which provide an intuitive yet rigorous means for defining and analyzing quantum processes, in place of the more traditional bra-ket notation. This expressive notation helps to clarify definitions and proofs, making them easier to read and understand, and encourages the use of equational (rather than calculational) reasoning.

The origins of CQM's graphical methods can be found in Penrose's tensor diagrams [30], although earlier graphical languages from physics (Feynman diagrams) and computer science (process charts) can be interpreted in these terms. Later, Joyal and Street [16] used category theory and topology to formalize these intuitive structures. More recently, the works of Selinger [33, 34] and Coecke, et al. [1, 7] have substantially tightened the connection between categorical methods and quantum information, in particular developing diagrammatic approaches positive maps and quantum-classical interaction, respectively. A thorough and self-contained introduction to this line of research can be found in the recent textbook [6].

[1] This condition allows the computations to be performed in polynomial time. The works [26, 11] go further, and prove an error term that is independent of the number of copies of the state.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

For a brief review of categorical quantum mechanics and the syntax of string diagrams used in our proofs, see Appendix A.

1.2 Statement of main result

In the three-player GHZ game, a referee chooses a random bit string $xyz \in \{0,1\}^3$ such that $x \oplus y \oplus z$=0, and distributes x, y, z to the three players (Alice, Bob, Charlie) respectively, who return numbers $a, b, c \in \{-1, 1\}$ respectively. The game is won if

$$abc \quad = \quad (-1)^{\neg(x \vee y \vee z)}. \tag{1}$$

(In other words, the game is won if the parity of the outputs is even and $xyz \neq 000$, or the parity of the outputs is odd and $xyz = 000$.) It is easy to prove that this game has no classical winning strategy. On the other hand, if Alice, Bob, and Charlie share the 3-qubit state

$$|G\rangle \quad = \quad \frac{1}{2\sqrt{2}} \quad \sum_{r+s+t \leq 1} |rst\rangle - \sum_{r+s+t \geq 2} |rst\rangle \Bigg) \Bigg(\tag{2}$$

and obtain their outputs by either performing an X-measurement $\{|+\rangle\langle+|, |-\rangle\langle-|\}$ on input 0 or the Z-measurement $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ on input 1, they win perfectly. This is known to be the only optimal strategy (up to local changes of basis) and therefore the GHZ game is rigid [20]. (An equivalent strategy, which is more conventional, is to use the state $|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ and X- and Y-measurements to win the GHZ game. We will use the state $|G\rangle$ instead for compatibility with previous work on rigidity. Note that $|G\rangle$ is equivalent under local unitary transformations to $|GHZ\rangle$.)

To extend this game to a self-test for the $|G\rangle^{\otimes N}$ state, we use a game modeled after [5]. The game requires the players to simulate playing the GHZ game N times. We give input to the rth player in the form of a pair (U_r, f_r) where $\{1, 2, \ldots, N\} \supseteq U_r \xrightarrow{f_r} \{0,1\}$ is a partial function assigning an "input" value for some subset of the game's "rounds" $1, \ldots, N$. The output given by such a player is a function $g_r \colon U_r \to \{0,1\}$ assigning a bit-valued "output" to each round. The game is won if the GHZ condition (1) is satisfied on all the rounds in $U_1 \cap U_2 \cap U_3$ for which the input string was even-parity.

Fortunately, it is not necessary to query the players on all possible subsets $U_r \subseteq \{0, 1, \ldots, N\}$ (which would involve an exponential number of inputs) — it is only necessary to query them on one- and two-element subsets. This yields the game $\overline{GHZ}_N$, which is formally defined in Figure 1 below. Our main result is the following. (See Proposition 10 below for a formal statement.)

Theorem 1. *The game $\overline{GHZ}_N$ is a self-test for the state $|G\rangle^{\otimes N}$, with error term $\mathcal{O}(N^4\sqrt{\varepsilon})$.*

In other words, if three devices succeed at the game $\overline{GHZ}_N$ with probability $1 - \varepsilon$, then the devices must contain a state that approximates the state $|G\rangle^{\otimes N}$ up to an error term of $\mathcal{O}(N^4\sqrt{\varepsilon})$. The proof proceeds by assuming that the players have such a high-performing strategy, and then using the measurements from that strategy to map their state isometrically to a state that is approximately of the form $|G\rangle \otimes L'$, where L' is some arbitrary tripartite "junk" state. This approach is a graphical translation of the method of many previous works on rigidity (in particular, [21] and our previous paper [17]).

1.3 Related works and further directions

In previous works on quantum rigidity, pictures are often used as an aid to a proof, but not as a proof itself. The only other rigidity paper that we know of which used rigorous graphical methods is the recent

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

Participants: Player 1 (Alice), Player 2 (Bob), Player 3 (Charlie), and a Referee
Parameter: $N \geq 1$ (the number of rounds)

1. The referee chooses a number $r \in \{0,1,2,3\}$ and a two distinct elements $i, j \in \{1,2,\ldots,N\}$ uniformly at random.
2. If $r = 0$, then the referee sets $U_1 = U_2 = U_3 = \{i\}$. Otherwise, he sets $U_r = \{i, j\}$ and sets each the other variables U_k (for $k \neq r$) to be equal to $\{i\}$.
3. The referee chooses values for $f_1(i), f_2(i), f_3(i) \in \{0,1\}$ uniformly at random under the constraint that $f_1(i) \oplus f_2(i) \oplus f_3(i) = 0$. Also, if $r > 0$, the referee chooses $f_r(j) \in \{0,1\}$ at random.
4. The referee sends each pair (U_i, f_i) to the ith player, respectively, who returns a function $g_i \colon U_i \to \{-1,1\}$.
5. The game is won if

$$g_1(i)g_2(i)g_3(i) \quad = \quad (-1)^{\neg(f_1(i) \vee f_2(i) \vee f_3(i))}. \tag{3}$$

Figure 1: The augmented GHZ game ($\overline{GHZ}_N$).

paper [12] which successfully used the concept of a group picture to prove rigidity for a new class of 2-player games. Group pictures are visual proofs of equations between elements of a finitely presented group. In the context of rigidity, group pictures construct approximate relations between products of sequences of operators, and as such they are a useful general tool for proving rigidity of 2-player games. An interesting further direction is to try to merge the formalism of [12] with the one given here in order to address general n-player games.

A natural next step is to explore cryptographic applications. Since GHZ states form the basis for the secret-sharing scheme of [15], it may be useful to see if the game $\overline{GHZ}_N$ can be used to create a new protocol for 3-party secret sharing using untrusted quantum devices.

2 Preliminaries

2.1 The augmented GHZ game

The game $\overline{GHZ}_N$ that we will use to self-test the state $|G\rangle^{\otimes N}$ from equation (2) is given in Figure 1. In this game, each player is requested to give outputs for either one or two round numbers (chosen from the set $\{1,2,\ldots,N\}$) given inputs for each round number. Both the inputs and the players' outputs are expressed as partial functions on the set $\{1,2,\ldots,N\}$. This game is modeled after [5].

The variable r determines the type of input given to each player. Note that in the case $r = 0$, there are $4N$ possible input combinations that the referee could give to the players (since there are N possible values for i, and 4 possible values for $(f_1(i), f_2(i), f_3(i))$) and for each of the values, $r = 1,2,3$, there are $8N(N-1)$ possible input combinations. Each valid input combination occurs with probability $\Omega(1/N^2)$.

We wish to describe the set of all possible quantum behaviors by the players Alice, Bob, and Charlie in $\overline{GHZ}_N$. In the definition that follows, we use the term *reflection* to mean an observable with values in $\{\pm 1\}$ (in other words, a Hermitian operator whose square is the identity). A *quantum strategy* for the

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

game $\overline{GHZ}_N$ game consists of the following data.

1. A unit vector $L \in A \otimes B \otimes C$, where A, B, C are finite-dimensional Hilbert spaces.
2. For each $i \in \{1,2,\ldots,N\}$, $b \in \{0,1\}$, and $W \in \{A,B,C\}$, a reflection

$$R^W_{i\to b} \tag{4}$$

   on W.

3. For each $i, j \in \{1,2,\ldots,N\}$, $b,c \in \{0,1\}$, and $W \in \{A,B,C\}$, two *commuting* reflections

$$R^W_{i\to b|j\to c} \quad \text{and} \quad R^W_{j\to c|i\to b} \tag{5}$$

   on W.

The spaces A, B, C denote the registers possessed by Alice, Bob, and Charlie, respectively. The vector L denotes the initial state that they share before the game begins. The reflections $R^W_{i\to b}$ describe their behavior on singleton rounds (specifically, on a singleton round the player measures his or her register W along the eigenspaces of $R^W_{i\to b}$, and reports either $+1$ or -1 for round i, appropriately). The reflections $R^W_{i\to b|j\to c}, R^W_{j\to c|i\to b}$ describe their behavior on non-singleton rounds (specifically, if the input to a player is the function $[i \to b, j \to c]$, then they measure along the eigenspaces of $R^W_{i\to b|j\to c}$ to obtain their output for round i and measure along the eigenspaces of $R^W_{j\to c|i\to b}$ to determine their output for round j). Note that since the reflections in (5) represent measurements that are carried out simultaneously by one of the players, we assume that these two reflections commute. (This assumption will be critical in our proof).

For any reflection Z and unit vector ψ on a finite-dimensional Hilbert space Q, if we measure ψ with Z then the probability of obtaining an output of -1 is precisely

$$[1 - \mathrm{Tr}(Z\psi\psi^*)]/2 \quad = \quad \|Z\psi - \psi\|^2/4 \tag{6}$$

We can use this fact to express the losing probabilities achieved by the players in terms of their strategy. If $r = 0$ and the players are queried for round i with inputs x, y, z, then their losing probability is precisely

$$\left\| R^A_{i\to x} R^B_{i\to y} R^C_{i\to z} L + (-1)^{x\vee y\vee z} L \right\|^2 /4. \tag{7}$$

If Alice is queried for round i with input x and round j with input x', and Bob and Charlie are queried for round i with inputs y, z respectively, then the losing probability is

$$\left\| R^A_{i\to x|j\to x'} R^B_{i\to y} R^C_{i\to z} L + (-1)^{x\vee y\vee z} L \right\|^2 /4. \tag{8}$$

Similar expressions hold for the case where Bob or Charlie is the party that receives two queries.

Note that the game $\overline{GHZ}_N$ is entirely symmetric between the three players Alice, Bob and Charlie. This means that, given any strategy $\left(L, \left\{R^W_{i\to b}\right., \left.\left\{R^W_{i\to b|j\to c}\right\}\right)$ for $\overline{GHZ}_N$, we can produce five additional strategies by choosing a nontrivial permutation $\sigma\colon \{1,2,3\} \to \{1,2,3\}$ and giving the pth players' subsystem and measurement strategy to the $\sigma(p)$th player, for each $p \in \{1,2,3\}$. We will make use of this symmetry in the proof that follows.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

2.2 Approximation chains

We make the following definition (see similar notation in [18, 4]). If $F, G\colon A \to B$ are linear maps, then

$$F \underset{\delta}{=} G \tag{9}$$

denotes the inequality $\|F-G\|_2 \leq \mathcal{O}(\delta)$, where $\|\cdot\|_2$ denotes the Frobenius norm and $\mathcal{O}$ denotes asymptotic big-O notation. (If F and G are vectors, then $\|F-G\|_2$ is simply the Euclidean distance $\|F-G\|$.)

We will use this notation especially in the case where F and G are processes represented by diagrams. Note that this relation is transitive: if $F \underset{\delta}{=} G$ and $G \underset{\delta}{=} H$, then $F \underset{\delta}{=} H$. Note also that if $J\colon B \to C$ is a linear map whose operator norm satisfies $\|J\|_\infty \leq \alpha$, then $F \underset{\delta}{=} G \Longrightarrow J \circ F \underset{\delta\alpha}{=} J \circ G$.

3 Rigidity of the augmented *GHZ* game

Throughout this section, suppose that

$$\left(L, \left\{R^A_{i\to a}\right\}, \left\{R^B_{i\to b}\right\}, \left\{R^C_{i\to c}\right\}, \left\{R^A_{i\to a|j\to a'}\right\}, \left\{R^B_{i\to b|j\to b'}\right\}, \left\{R^C_{i\to c|j\to c'}\right\}\right) \tag{10}$$

is a quantum strategy for the $\overline{GHZ}_N$ game which wins with probability $1-\varepsilon$. It is helpful to introduce some redundant notation for this strategy: for any $W \in \{A,B,C\}$ and $i \in \{1,2,\ldots,N\}$ let

$$X'_{W,i} = R^W_{i\mapsto 0}, \tag{11}$$

$$Z'_{W,i} = R^W_{i\mapsto 1}, \tag{12}$$

The reason for this notation is that we intend to show that the operators $R^W_{i\mapsto 1}, R^W_{i\mapsto 1}$ approximate the behavior of the X and Z measurements in the optimal GHZ strategy (see the beginning of subsection 1.2). Similarly, let $X'_{W,i|j\to 1} = R^W_{i\to 0|j\to 1}$ and $Z'_{W,j|i\to 0} = R^W_{j\to 1|i\to 0}$.

We will drop the subscript W from this notation when it is clear from the context.

3.1 Initial steps

Our first goal is to prove approximate commutativity and anticommutativity relations for the operators $X'_{W,i}, Z'_{W,i}$.

Since the losing probability for our chosen strategy is ε, and each input string occurs with probability $\Omega(N^2)$, we can conclude that the probability of losing on any particular input combination is $\mathcal{O}(N^2\varepsilon)$. By the discussion of expressions (7) and (8) above, we therefore have the following for any $i \in \{1,2,\ldots,N\}$:

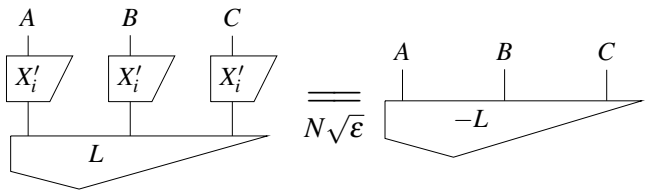
$$\begin{aligned}
\left\|(I + X'_{A,i}X'_{B,i}X'_{C,i})|L\rangle\right\|^2 &\leq \mathcal{O}(N^2\varepsilon),\\
\left\|(I - Z'_{A,i}Z'_{B,i}X'_{C,i})|L\rangle\right\|^2 &\leq \mathcal{O}(N^2\varepsilon),\\
\left\|(I - X'_{A,i}Z'_{B,i}Z'_{C,i})|L\rangle\right\|^2 &\leq \mathcal{O}(N^2\varepsilon),\\
\left\|(I - Z'_{A,i}X'_{B,i}Z'_{C,i})|L\rangle\right\|^2 &\leq \mathcal{O}(N^2\varepsilon),
\end{aligned} \tag{13}$$

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

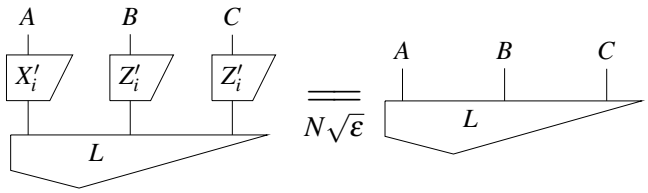
Additionally, if we take any of the four inequalities above, and replace any one of the operators $X'_{W,i}$ with $X'_{W,i|j\to t}$, where $j \neq i$ and $t \in \{0,1\}$, the inequality remains true, and likewise if we replace any $Z'_{W,i}$ with $Z'_{W,i|j\to t}$, the inequality remains true.

We can translate the inequalities above into graphical form.

Proposition 2. *The following inequalities hold.*



(14)

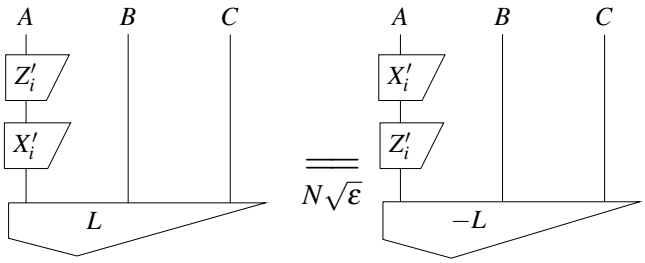


(15)

And, inequality (15) also holds for any permutation of the letters A, B, C.

We will use Proposition 2 to prove the following assertion.

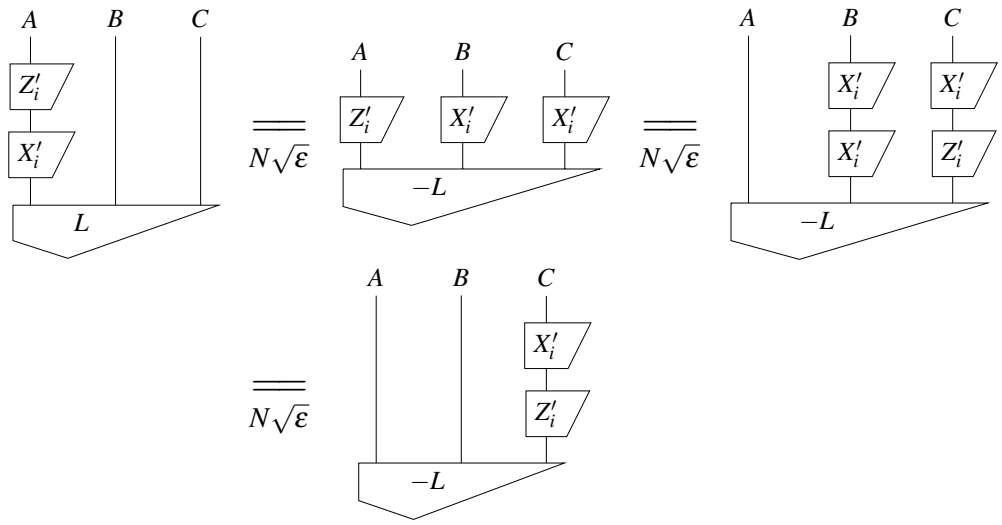
Proposition 3 (Approximate anti-commutativity)**.** *For any i, the following inequality holds:*



(16)

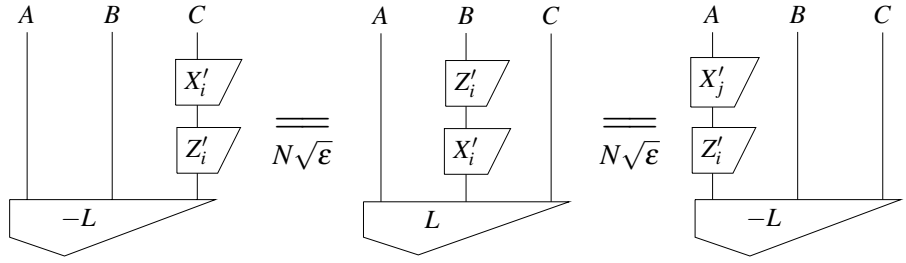
Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

Proof. Repeatedly applying Proposition 2,



(17)

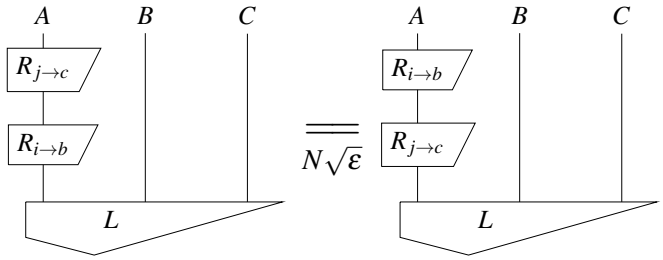
(Here we have used the fact that all of the unitary maps in these diagrams are self-inverse.) Applying the same steps symmetrically across the wires A, B, C,



(18)

as desired. □

Proposition 4 (Approximate commutativity)**.** *The following equation holds for any distinct $i, j \in \{1, 2, \ldots, N\}$ and any bits $b, c \in \{0, 1\}$:*



The proof of Proposition 4 is given in appendix C, and is based on the fact that the related reflections $R_{i\to b|j\to c}$ and $R_{j\to c|i\to b}$ commute.

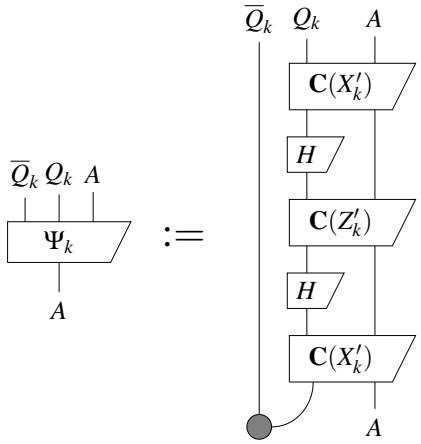
Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

3.2 Isometries

Next we define the local isometries which will relate the state L to the ideal state $|GHZ\rangle^{\otimes N}$. Roughly speaking, for each $k \in \{1,2,\ldots,N\}$ and $W \in \{A,B,C\}$, we will construct an isometry $\Psi_{W,k}$ which approximately "locates" a qubit with the ith players' system, and swaps it out onto a qubit register $Q \cong \mathbb{C}^2$. We then apply these isometries in order, for $1,2,\ldots,N$, the system W. We will use approximate commutativity to show that the different isometries $\Psi_{W,k}$ do not interfere much with one another when applied in sequence. Our approach borrows from previous works on rigidity and uses similar notation to that of the non-graphical rigidity proof in our previous paper [17].

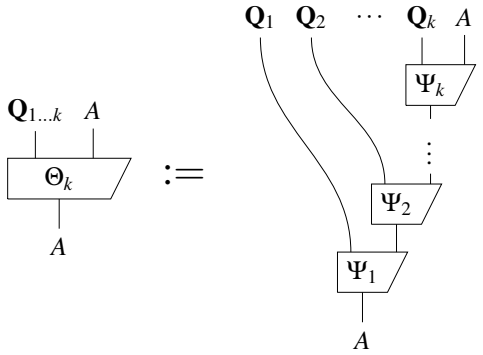
In the following, we use controlled unitary gates, $\mathbf{C}(U)$, which are defined and discussed in Appendix B. Let $H\colon \mathbb{C}^2 \to \mathbb{C}^2$ denote the Hadamard gate $[|0\rangle \mapsto |+\rangle, |1\rangle \mapsto |-\rangle]$. We define isometries $\Psi_{A,k}$ on the system A which involve preparing a Bell state (denoted by a gray node) and then performing an approximate "swap" procedure between one half of the Bell state and A. (This is based on [21].) Then we define an isometry $\Theta_{A,k}$ on A which chains together the swapping maps $\Psi_{A,1},\ldots,\Psi_{A,k}$.

Definition 5 (Swapping maps)**.** *For each $k \in \{1,2,\ldots,N\}$, let Q_k and $\overline{Q}_k$ denote qubit registers, and define an isometry $\Psi_{A,k}$ as follows (suppressing the label A when it is not necessary):*



(19)

Let $\mathbf{Q}_k = \overline{Q}_k \otimes Q_k$ and $\mathbf{Q}_{1\ldots k} = \mathbf{Q}_1 \otimes \cdots \otimes \mathbf{Q}_k$. Define an isometry $\Theta_{k,A}$ by



(20)

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

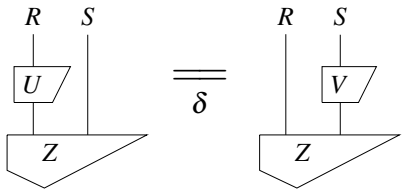
Let $Q_{N+1},\dots,Q_{3N},\overline{Q}_{N+1},\dots,\overline{Q}_{3N}$ be qubit registers, and define $\Psi_{B,k}$ analogously as an isometry from B to $B\otimes\overline{Q}_{N+k}\otimes Q_{N+k}$. Define $\Psi_{C,k}$ analogously as an isometry from C to $C\otimes\overline{Q}_{2N+k}\otimes Q_{2N+k}$. Define composite maps $\Theta_{B,k}$ and $\Theta_{C,k}$ similarly in terms of $\Psi_{B,k}$ and $\Psi_{C,k}$.

3.3 Commutativity properties

We now investigate some approximate (anti-)commutativity relationships between the Pauli operators on Q, the reflection strategies for A, B and C, and the isometries defined in the last section.

We begin with the following definition and lemma, which are crucial.

Definition 6. *Let R,S be registers and let $Z\in R\otimes S$ be a unit vector. If a unitary map $U\colon R\to R$ is such that there exists another unitary map $V\colon S\to S$ satisfying*



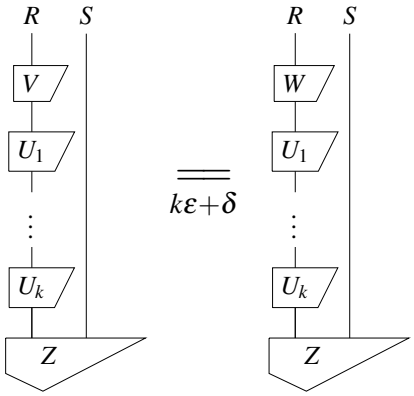
(21)

the we say that U can be pushed through *Z* with error term *δ*.

Lemma 7 (Push Lemma)**.** *Suppose that R,S are registers, $Z\in R\otimes S$ is a unit vector, and $V,W,U_1,U_2,\dots,U_k$ are unitary operators on R such that*

1. *Each map U_i can be pushed through Z with error term ε, and*
2. *The approximate equality $(V\otimes I_S)L \underset{\delta}{=} (W\otimes I_S)L$ holds.*

Then,



(22)

The Push Lemma follows from an easy inductive argument, and is given in the appendix. Note that by Proposition 2, for any k we have

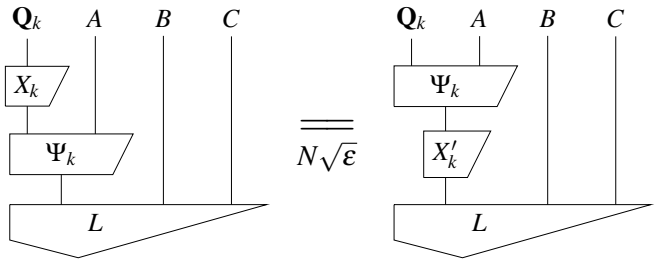
$$X'_{A,k}L \underset{N\sqrt{\varepsilon}}{=} (-X'_{B,k}\otimes X'_{C,k})L \tag{23}$$

$$Z'_{A,k}L \underset{N\sqrt{\varepsilon}}{=} (Z'_{B,k}\otimes X'_{C,k})L, \tag{24}$$

and so all of the maps $X'_{\cdot,k}$ and $Z'_{\cdot,k}$ can be pushed through L with error term $N\sqrt{\varepsilon}$. This fact underlies the proofs of the next two results, which are proved in the appendix using a combination of the Push Lemma, and the approximate commutativity and anti-commutativity properties of maps $X'_{\cdot,k}$ and $Z'_{\cdot,k}$.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

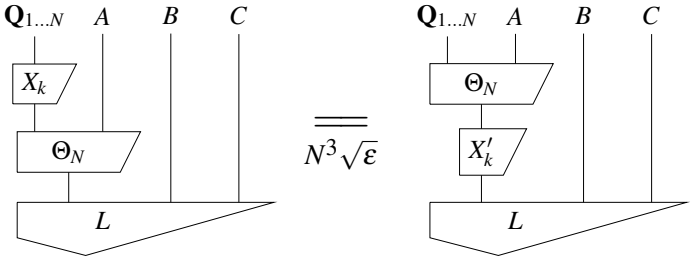
Proposition 8. *For $k \in \{1,2,\ldots,N\}$, let $X_{A,k}$ and $Z_{A,k}$ denote the Pauli operators on Q_k. Then,*



(25)

and similarly for Z'_k. Likewise, define $\{X_{B,k}, Z_{B,k}\}$ to be the Pauli operators on Q_{N+k} and define $\{X_{C,k}, Z_{C,k}\}$ to be the Pauli operators on Q_{2N+k}. Analogous statements hold for $\Psi_{k,B}$ and $\Psi_{k,C}$.

Proposition 9. *For any $k \in \{1,2,\ldots,N\}$,*



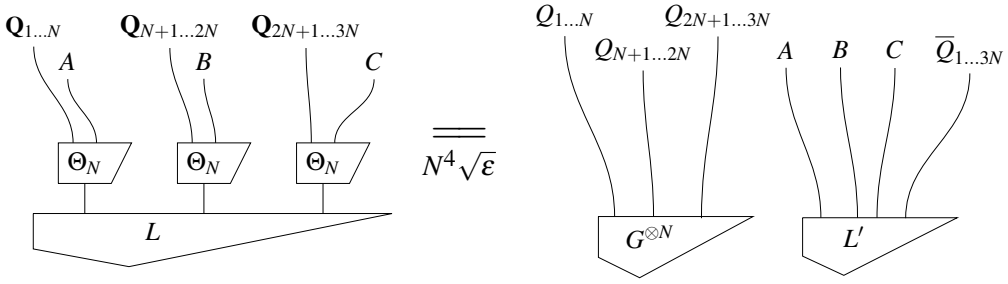
(26)

and similarly for Z'_k. Analogous statements hold for $\Theta_{B,k}$ and $\Theta_{C,k}$.

3.4 Rigidity

We are now ready to state and prove our main result.

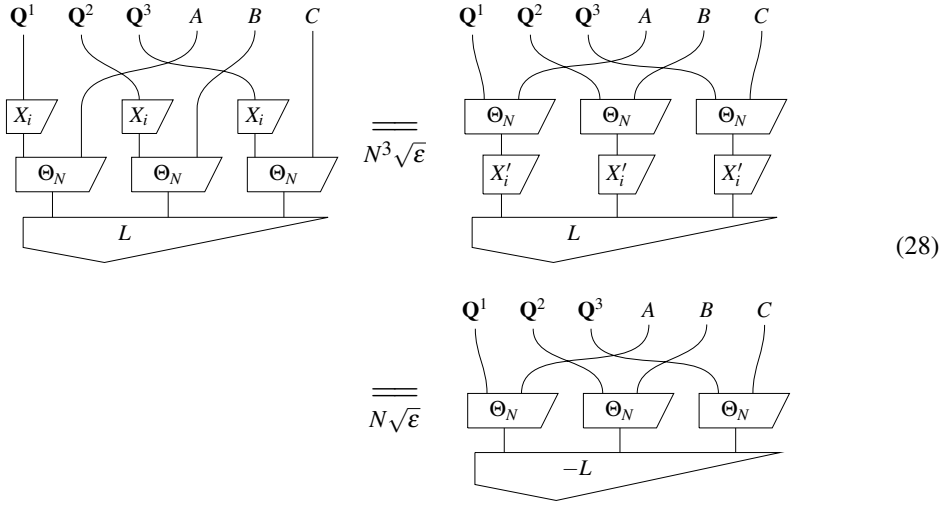
Proposition 10 (Rigidity)**.** *Let $\Theta_{A,N}, \Theta_{B,N}, \Theta_{C,N}$ be the isometries from Definition 5. Then, there is some state L' on $A \otimes B \otimes C \otimes \overline{Q}_{1\ldots 3N}$ such that*



(27)

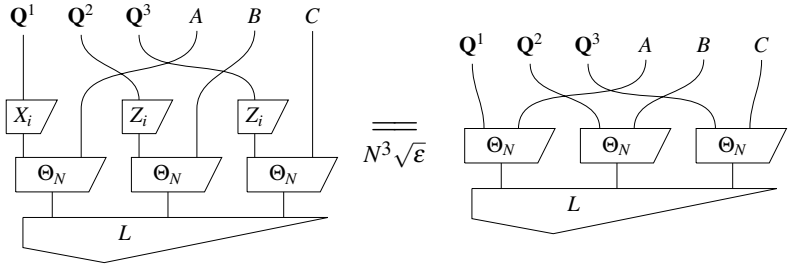
Proof. In the following, we write $\mathbf{Q}^1 := \mathbf{Q}_{1\ldots N}$, $\mathbf{Q}^2 := \mathbf{Q}_{(N+1)\ldots 2N}$ and $\mathbf{Q}^3 := \mathbf{Q}_{(2N+1)\ldots 3N}$ in order to conserve space. By application of Props. 2 and 9 we have,

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.



(28)

(Here, X_i is used to denote the X-Pauli operator on either Q_i, Q_{N+i}, or Q_{2N+i} depending on which wire it is applied to.) Similarly we obtain that



(29)

where the last relation also holds for any permutation of the labels (X_i, Z_i, Z_i) on the left side of the equation.

The commuting reflection operators $X \otimes Z \otimes Z$, $Z \otimes X \otimes Z$, and $Z \otimes Z \otimes X$ on $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ have a common orthonormal eigenbasis $G = G_0, G_1, \ldots, G_7$, in which G_0 is the only eigenvector that has eigenvalue $(+1)$ for all three operators. We can express $\Theta_A^N \otimes \Theta_B^N \otimes \Theta_c^N |L\rangle$ using this basis as

$$\Theta_A^N \otimes \Theta_B^N \otimes \Theta_c^N |L\rangle = \sum_{v_1,\cdots,v_N \in \{G_0,\ldots,G_7\}} |v_1\rangle \otimes \cdots \otimes |v_N\rangle \otimes |L'_{\mathbf{v}}\rangle. \tag{30}$$

where $L'_{\mathbf{v}} \in A \otimes B \otimes C \otimes \overline{Q}_{1\ldots 3N}$. For every term in the sum on the right except the one indexed by $G_0^{\otimes N}$, there is an operator of the form $X_{A,i} \otimes Z_{B,i} \otimes Z_{C,i}$, $Z_{A,i} \otimes X_{B,i} \otimes Z_{C,i}$, or $Z_{A,i} \otimes Z_{B,i} \otimes X_{C,i}$ which negates it. By equation (29) above, the total length of all the terms negated by any one particular gate of this form is $\mathcal{O}(N^3\sqrt{\varepsilon})$, and so the total length of all terms in (30) other than the $G_0^{\otimes N}$ term is $\mathcal{O}(N^4\sqrt{\varepsilon})$, as desired. □

We note that our proofs generalize in a straightforward manner to a proof of self-testing for an arbitrary number of copies of a k-GHZ state, for any integer $k > 3$. The graphical method seems generally

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

well-suited to proving parallel self-testing for stabilizer states, including graph states [21]. We leave possible generalizations to future work.

References

[1] Samson Abramsky & Bob Coecke (2004): *A categorical semantics of quantum protocols*. In: *Logic in computer science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, IEEE, pp. 415–425.

[2] John Baez & Mike Stay (2010): *Physics, topology, logic and computation: a Rosetta Stone*. In: *New structures for physics*, Springer, pp. 95–172.

[3] Charles H Bennett & Gilles Brassard (2014): *Quantum cryptography: Public key distribution and coin tossing. Theor. Comput. Sci.* 560(P1), pp. 7–11.

[4] Spencer Breiner, Carl A Miller & Neil J Ross (2017): *Graphical Methods in Device-Independent Quantum Cryptography*. *arXiv preprint arXiv:1705.09213*.

[5] Rui Chao, Ben W Reichardt, Chris Sutherland & Thomas Vidick (2016): *Test for a large amount of entanglement, using few measurements*. *arXiv preprint arXiv:1610.00771*.

[6] Bob Coecke & Aleks Kissinger (2017): *Picturing quantum processes*. Cambridge University Press.

[7] Bob Coecke & Dusko Pavlovic (2006): *Quantum measurements without sums*. *arXiv preprint quant-ph/0608035*.

[8] Andrea Coladangelo (2017): *Parallel self-testing of (tilted) EPR pairs via copies of (tilted) CHSH and the magic square game*. *Quantum Information and Computation* 17(9-10), pp. 831–865.

[9] Andrea Coladangelo (2018): *A generalization of the CHSH inequality self-testing maximally entangled states of any local dimension*. *arXiv preprint arXiv:1803.05904*.

[10] Andrea Coladangelo, Koon Tong Goh & Valerio Scarani (2017): *All pure bipartite entangled states can be self-tested*. *Nature communications* 8, p. 15485.

[11] Andrea Coladangelo, Alex Grilo, Stacey Jeffery & Thomas Vidick (2017): *Verifier-on-a-Leash: new schemes for verifiable delegated quantum computation, with quasilinear resources*. *arXiv preprint arXiv:1708.07359*.

[12] Andrea Coladangelo & Jalex Stark (2017): *Robust self-testing for linear constraint system games*. *arXiv preprint arXiv:1709.09267*.

[13] Matthew Coudron & Anand Natarajan (2016): *The parallel-repeated magic square game is rigid*. *arXiv preprint arXiv:1609.06306*.

[14] Matteo Fadel (2017): *Self-testing Dicke states*.

[15] Mark Hillery, Vladimír Bužek & André Berthiaume (1999): *Quantum secret sharing*. *Physical Review A* 59(3), p. 1829.

[16] André Joyal & Ross Street (1991): *The geometry of tensor calculus, I*. *Advances in mathematics* 88(1), pp. 55–112.

[17] Amir Kalev & Carl A Miller (2017): *Rigidity of the magic pentagram game*. *Quantum Science and Technology* 3(1), p. 015002.

[18] Aleks Kissinger, Sean Tull & Bas Westerbaan (2017): *Picture-perfect Quantum Key Distribution*. ArXiv:1704.08668.

[19] Dominic Mayers & Andrew Yao (1998): *Quantum cryptography with imperfect apparatus*. In: *Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on*, IEEE, pp. 503–509.

[20] Matthew McKague (2011): *Self-testing graph states*. In: *Conference on Quantum Computation, Communication, and Cryptography*, Springer, pp. 104–120.

[21] Matthew McKague (2016): *Interactive Proofs for BQP via Self-Tested Graph States*. *Theory of Computing* 12(3), pp. 1–42.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

[22] Matthew McKague (2016): *Self-testing in parallel*. *New Journal of Physics* 18(4), p. 045013.

[23] Matthew McKague (2017): *Self-testing in parallel with CHSH*. *Quantum* 1, p. 1.

[24] Matthew McKague, Tzyh Haur Yang & Valerio Scarani (2012): *Robust self-testing of the singlet*. *Journal of Physics A: Mathematical and Theoretical* 45(45), p. 455304.

[25] Carl A. Miller & Yaoyun Shi (2013): *Optimal Robust Self-Testing by Binary Nonlocal XOR Games*. In: *8th Conference on the Theory of Quantum Computation, Communication and Cryptography*, pp. 264–272.

[26] Anand Natarajan & Thomas Vidick (2017): *A Quantum Linearity Test for Robustly Verifying Entanglement*. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, ACM, New York, NY, USA, pp. 1003–1015, doi:10.1145/3055399.3055468. Available at `http://doi.acm.org/10.1145/3055399.3055468`.

[27] Anand Natarajan & Thomas Vidick (2018): *Low-degree testing for quantum states*. *arXiv preprint arXiv:1801.03821*.

[28] Dimiter Ostrev (2016): *The structure of nearly-optimal quantum strategies for the CHSH (n) XOR games*. *Quantum Information & Computation* 16(13-14), pp. 1191–1211.

[29] Károly F. Pál, Tamás Vértesi & Miguel Navascués (2014): *Device-independent tomography of multipartite quantum states*. *Phys. Rev. A* 90, p. 042340, doi:10.1103/PhysRevA.90.042340. Available at `https://link.aps.org/doi/10.1103/PhysRevA.90.042340`.

[30] Roger Penrose (1971): *Applications of negative dimensional tensors*. *Combinatorial mathematics and its applications* 1, pp. 221–244.

[31] Sandu Popescu & Daniel Rohrlich (1992): *Which states violate Bell's inequality maximally?* *Physics Letters A* 169(6), pp. 411–414.

[32] Ben W Reichardt, Falk Unger & Umesh Vazirani (2013): *Classical command of quantum systems*. *Nature* 496(7446), p. 456.

[33] Peter Selinger (2004): *Towards a quantum programming language*. *Mathematical Structures in Computer Science* 14(4), pp. 527–586.

[34] Peter Selinger (2007): *Dagger compact closed categories and completely positive maps*. *Electronic Notes in Theoretical computer science* 170, pp. 139–163.

[35] Peter Selinger (2010): *A survey of graphical languages for monoidal categories*. In: *New structures for physics*, Springer, pp. 289–355.

[36] Ivan Šupić, Andrea Coladangelo, Remigiusz Augusiak & Antonio Acín (2017): *A simple approach to self-testing multipartite entangled states*. ArXiv:1707.06534.

[37] Xingyao Wu (2016): *Self-Testing: Walking on the Boundary of the Quantum Set*. Ph.D. thesis, National University of Singapore.

[38] Xingyao Wu, Yu Cai, Tzyh Haur Yang, Huy Nguyen Le, Jean-Daniel Bancal & Valerio Scarani (2014): *Robust self-testing of the three-qubit W state*. *Physical Review A* 90(4), p. 042339.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

Appendix A Categorical Quantum Mechanics

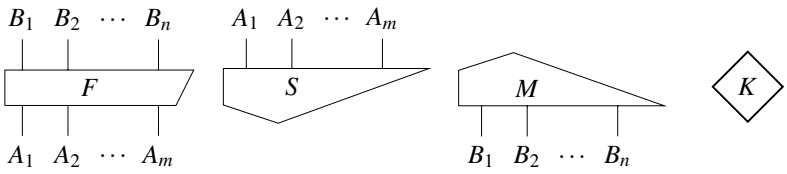
In this appendix we will briefly review some of the standard machinery of categorical quantum mechanics. For a thorough introduction to the topic, see [6].

A.1 Symmetric Monoidal Categories

The formal context for categorical quantum mechanics is that of *symmetric monoidal categories*. We develop the terminology in stages.

A *category* is a mathematical structure representing a universe of (possible) processes $F, G, H, \ldots$; each process has a typed input and output, often indicated by writing $F : A \to B$. The fundamental structure in a category is serial composition; whenever the output type of $F : A \to B$ matches the input type of $G : A \to C$, we may form a composite process $F \circ G : A \to C$.

A *monoidal* category generalizes the structure of an ordinary category to allows for multi-partite processes; here the fundamental object of study is a process $F : A_1 \otimes \ldots \otimes A_m \to B_1 \otimes \ldots \otimes B_n$, which is represented as a black box with m labeled inputs and n labeled outputs. (Either of the numbers m, n may be zero.) Diagramatically, we represent these classes as follows:

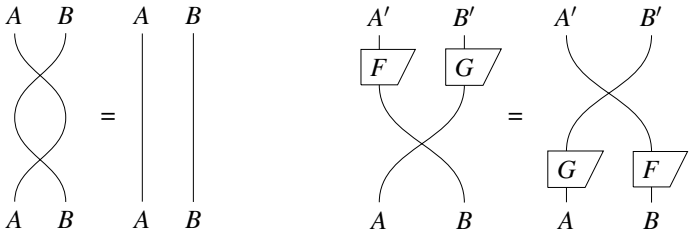


(31)

The categorical structure of serial composition is represented diagrammatically by matching the output wires of one process to the inputs of another, so long as the types match up. So, above, F can be pre-composed with S and post-composed with M to yield a scalar $M \circ F \circ S$ or, in Dirac notation, $\langle M|F|S\rangle$.

Along with multi-partite states and processes, monoidal structure also introduces an operation of parallel composition on processes: given $F : A \to B$ and $G : A' \to B'$, we can produce a parallel process $F \otimes G : A \otimes A' \to B \otimes B'$, depicted graphically by side-by-side juxtaposition. More generally, using parallel composition with identity processes (represented by bare wires), we can compose processes in which only some inputs and outputs match. Note that we will often suppress wire labels in complicated diagrams, as the labels are implicitly determined by the boxes they feed.

Finally, a *symmetric* structure on a monoidal category allow for additional flexibility in how wires can be manipulated. A symmetry allows us to permute the ordering of strings, and is represented diagrammatically by crossing wires (called a *twist*). Formally, the twist is axiomatized terms of intuitive diagrammatic equations:



(32)

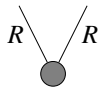
Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

A.2 Quantum states and processes

To apply the above approach to quantum mechanics, we work within the category **Hilb** of finite-dimensional Hilbert spaces over $\mathbb{C}$. In this category, the types are finite-dimensional Hilbert spaces (i.e., vector spaces of $\mathbb{C}$ with semi-linear inner product) each equipped with a fixed orthonormal basis, and the processes are $\mathbb{C}$-linear maps between such vector spaces. For example, if $A_1,\ldots,A_m,B_1,\ldots,B_n$ are finite-dimensional Hilbert spaces, then the diagrams in display box (31) above represent, respectively, a linear map $F\colon A_1\otimes\ldots\otimes A_m\to B_1\otimes\ldots\otimes B_n$, a vector $S\in A_1\otimes\ldots\otimes A_m$, a linear map $M\colon B_1\otimes\ldots\otimes B_n\to\mathbb{C}$, and a scalar $K\in\mathbb{C}$. Serial composition of diagrams simply represents composition of functions — for example, the composition of S,F and M is simply the scalar $M(F(S))\in\mathbb{C}$. It is elementary to show this category is a symmetric monoidal category (with the tensor product as its monoidal operation).

For our purposes, a *state* is a vector in a Hilbert space (i.e., a process with no inputs) whose norm is equal to one. A *unitary process* $F\colon A\to B$ is a linear map that satisfies $FF^*=I_B$ and $F^*F=I_A$. (Also, a linear map $G\colon A\to B$ that satisfies the single condition $G^*G=I_A$ is called an *isometry*.) With these definitions, we will be able to express quantum states and processes as diagrams like the ones in (31) and (32) above.

We note that while symmetric monoidal categories are sufficient to handle the book-keeping needed in our proofs, it is only a fragment of the full CQM theory. Further development introduces compact closed structures (trace, transpose, state-process duality), dagger structures (adjoint, conjugate), Frobenius structures (orthonormal basis, classical-quantum interaction) and Hopf structures (complementary bases, ZX-calculus). For our purposes, we need only one additional visual definition, which is the *Bell state*. In this paper we use a gray node with two wires of the same type R,



(33)

to denote the unit vector

$$\left(\frac{1}{\sqrt{\dim R}}\right)\sum_e e\otimes e \quad\in\quad R\otimes R, \tag{34}$$

where the sum is taken over the standard basis of R. If $R\cong\mathbb{C}^2$, we denote this state symbolically by Φ^+.

Appendix B Controlled Unitaries

Our proof makes substantial uses of controlled unitary operations. This appendix collects key facts about controlled operations which will simplify our main proof.

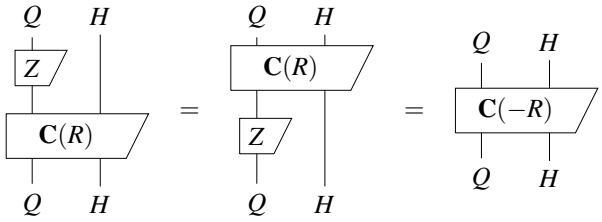
Definition 11 (Controlled Unitary)**.** *Let $Q\cong\mathbb{C}^2$ denote a qubit register with a fixed computational basis $\{|0\rangle,|1\rangle\}$, and suppose $U\colon H\to H$ is a unitary operation. The associated* controlled unitary $\mathbf{C}(U)\colon Q\otimes H\to Q\otimes H$ *is defined by*

$$\mathbf{C}(U) \quad=\quad |0\rangle\langle 0|\otimes I_H+|1\rangle\langle 1|\otimes U. \tag{35}$$

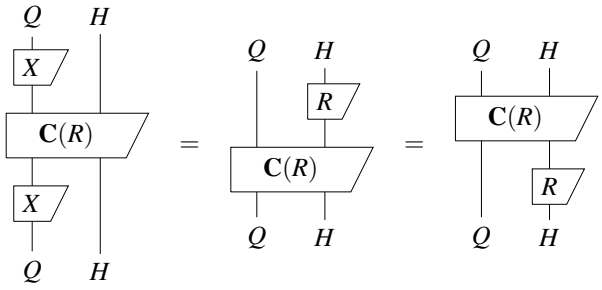
The next lemma describes some (anti-)commutativity properties between controlled unitaries and Pauli operators X and Z. The proofs follow directly from the definition.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

Lemma 12. *For any reflection $R : H \to H$ we have the following equations*



(36)



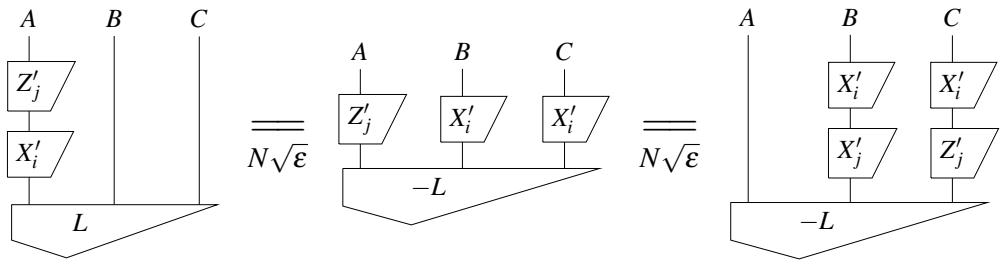
(37)

Appendix C Supporting Proofs

In this appendix we provide the proofs for propositions from the main text.

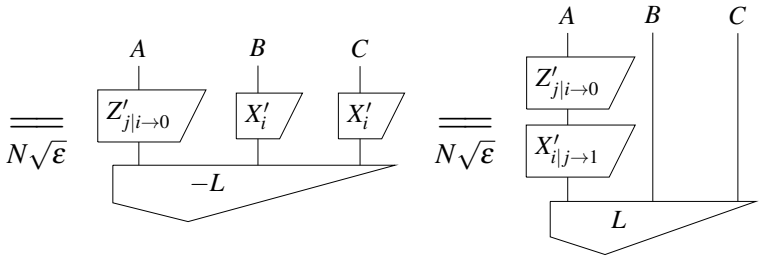
C.1 Proof of Proposition 4

We give the proof for $b = 0, c = 1$; the other cases are analogous. Using inequalities (13) and their variants, we have

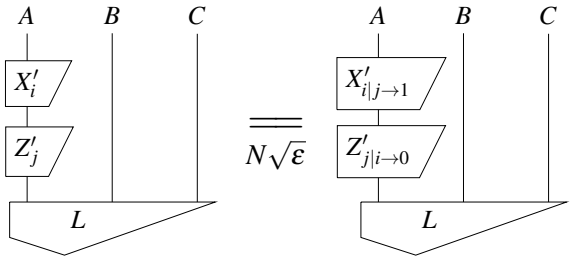


(38)

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.



A similar sequence shows that

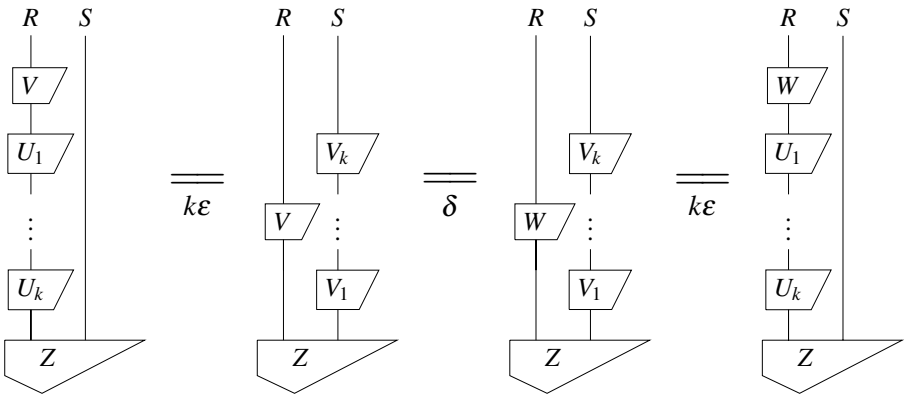


(39)

Since $X'_{A,i|j\to 1} = R^A_{i\to 0|j\to 1}$ and $Z'_{A,j|i\to 0} := R^A_{j\to 1|i\to 0}$ are assumed to commute, the result follows.

C.2 Proof of Lemma 7

Applying the push condition inductively, we have the following for some unitary operators $V_1, \ldots, V_k$:

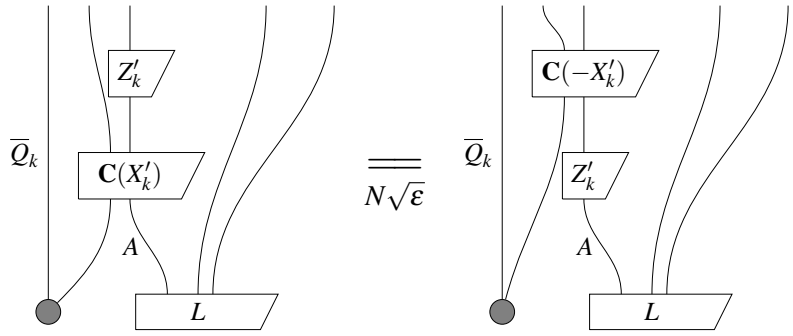


(40)

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

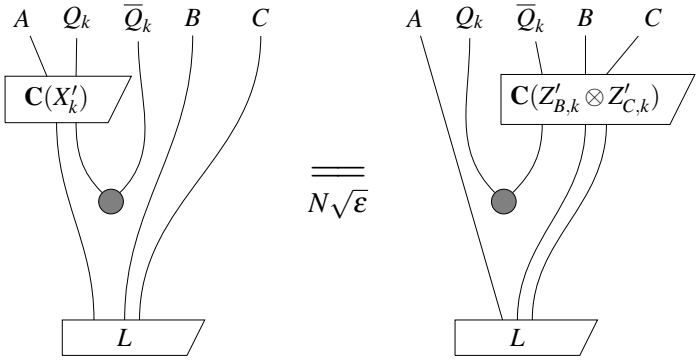
C.3 Proof of Proposition 8

We begin with two observations. First, the approximate anti-commutativity in Proposition 3 implies the following approximate anti-commutativity property for the controlled-X'_k gate (by superposition):



(41)

Secondly, the controlled operator $\mathbf{C}(X'_{A,k})$ on $Q_k \otimes A$ can be approximately pushed through the state $\Phi^+ \otimes L$ like so:

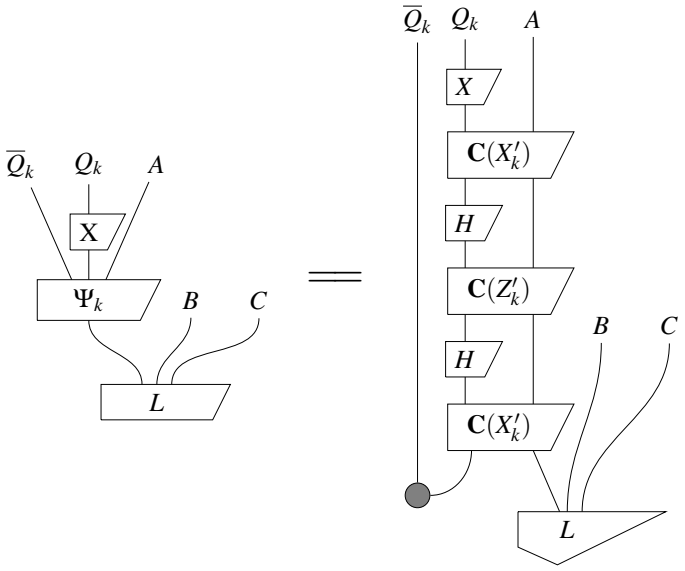


(42)

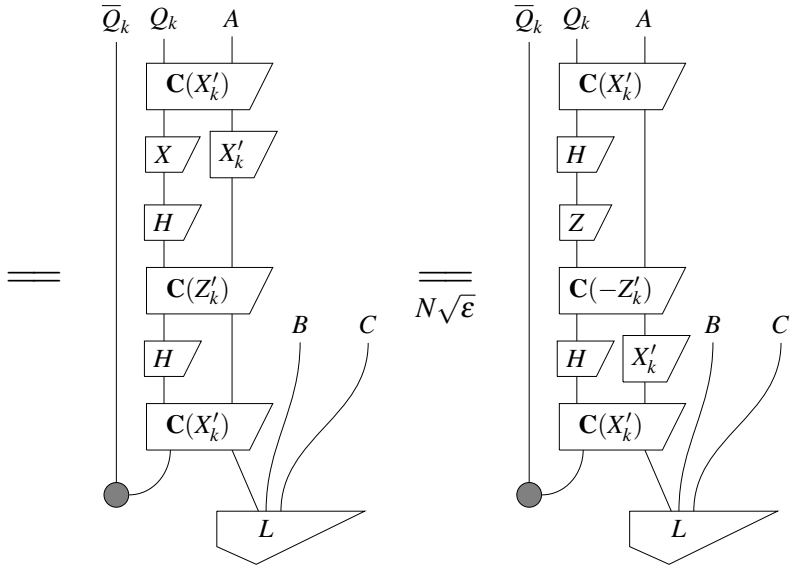
And similarly for $Z'_{A,k}$. The Hadamard operator $[H \otimes I_A]$ on $Q_k \otimes A$ can be exactly pushed through $\Phi^+ \otimes L$ (by merely applying H to $\overline{Q}_k$). This fact allows free application of the Push Lemma (Lemma 7). We have the following, in which we exploit approximate anti-commutativity, the Lemma 7, and the rules for

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

controlled unitaries from Appendix B.

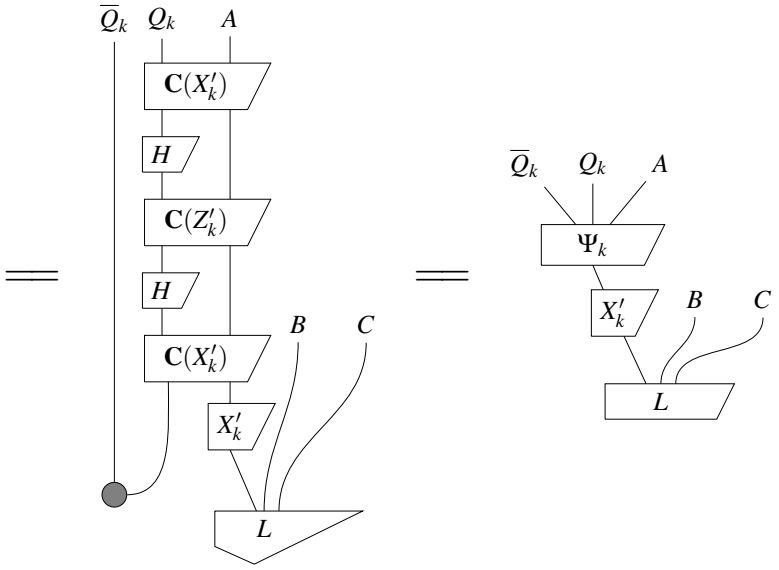


(43)



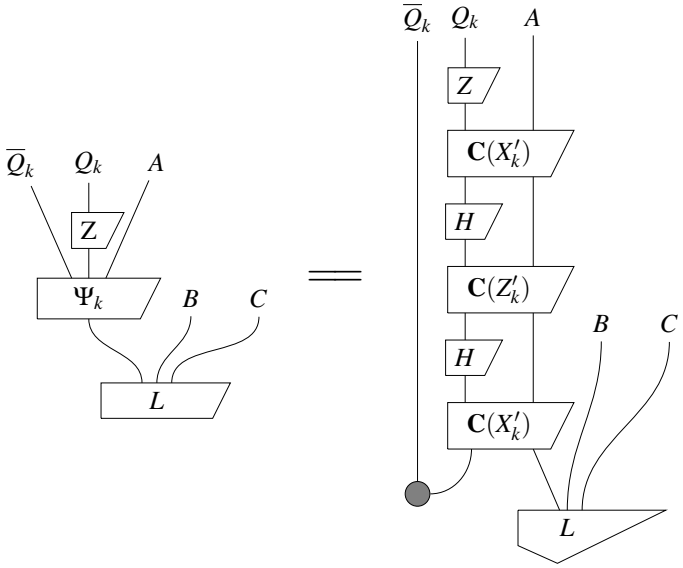
(44)

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.



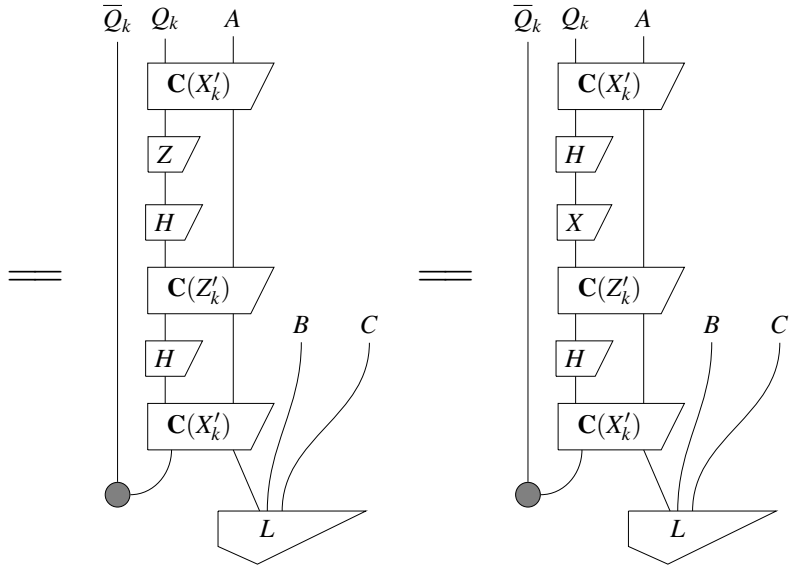
(45)

Similarly,

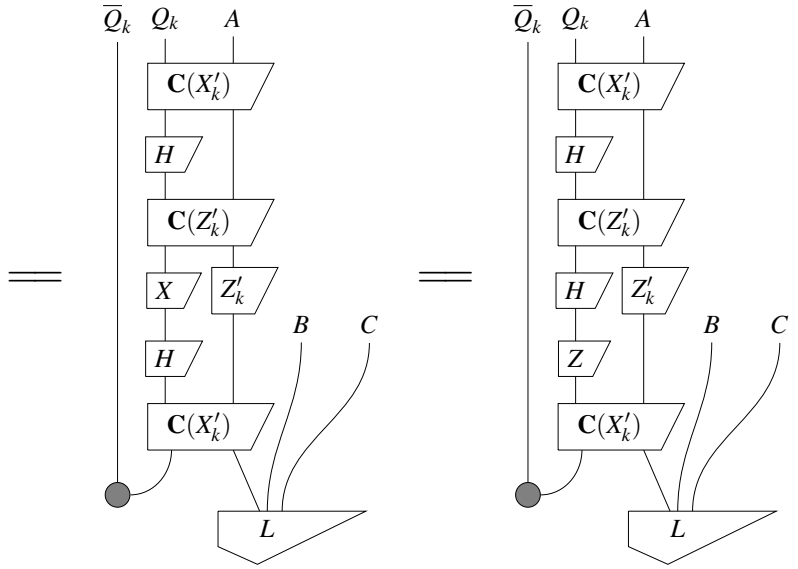


(46)

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

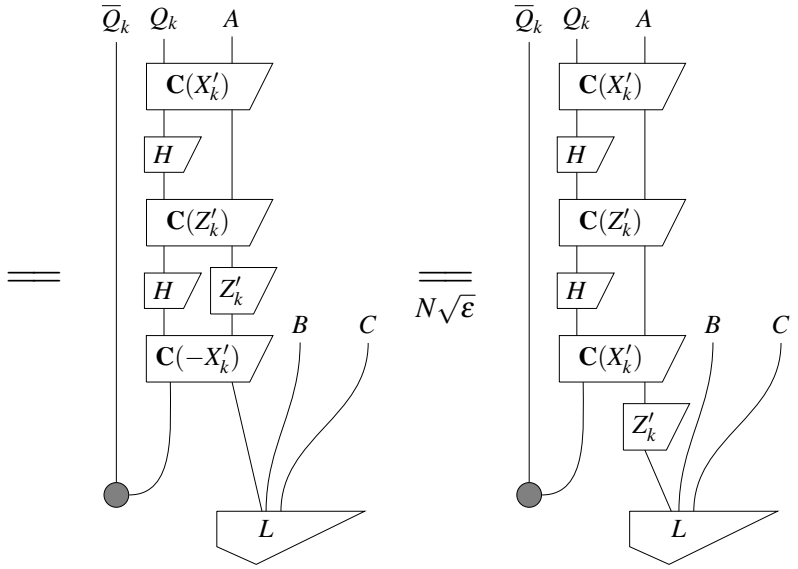


(47)

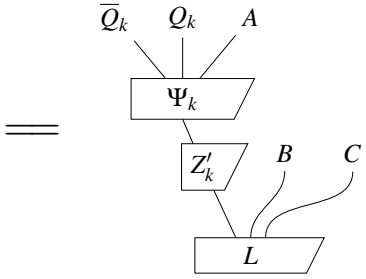


(48)

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.



(49)



(50)

as desired. This completes the proof.

C.4 Proof of Proposition 9

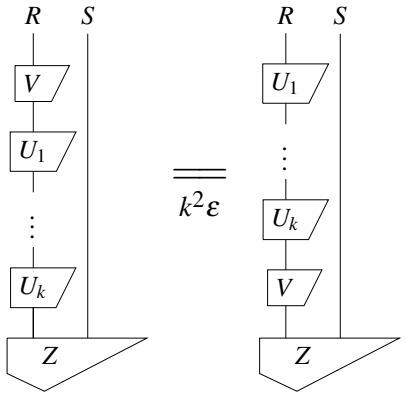
We begin with the following lemma. It is similar to the Push Lemma (Lemma 7) but it specifically addresses commutativity.

Lemma 13. *Suppose that R, S are registers, $Z \in R \otimes S$ is a unit vector, and $V, U_1, U_2, \ldots, U_k$ are unitary operators on R such that*

1. *Each map U_i can be pushed through Z with error term ε, and*
2. *The approximate equality $(VU_i \otimes I_S)L \underset{\varepsilon}{=} (U_i V \otimes I_S)L$ holds for all i.*

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

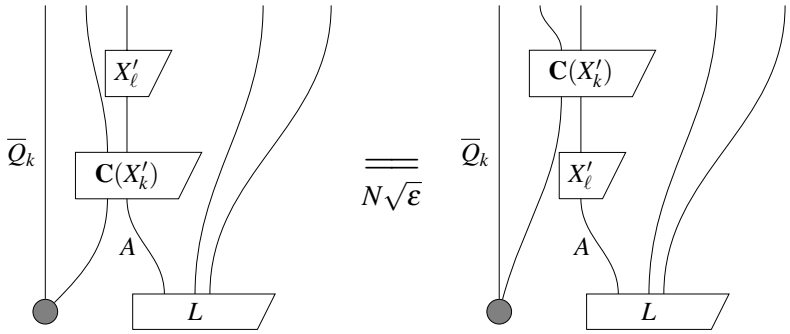
Then,



(51)

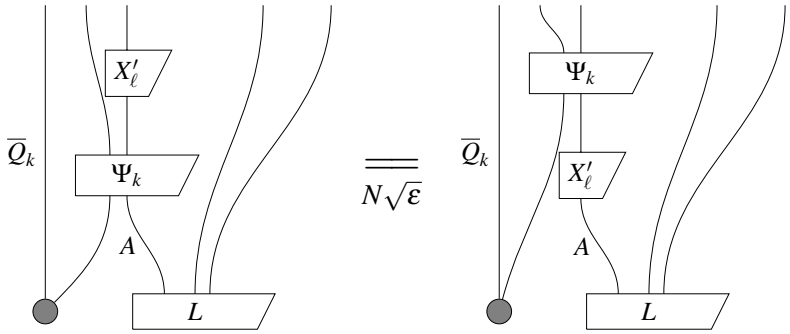
Proof. This follows easily by k applications of Lemma 7. □

By Proposition 4, for any $j \neq k$, we have



(52)

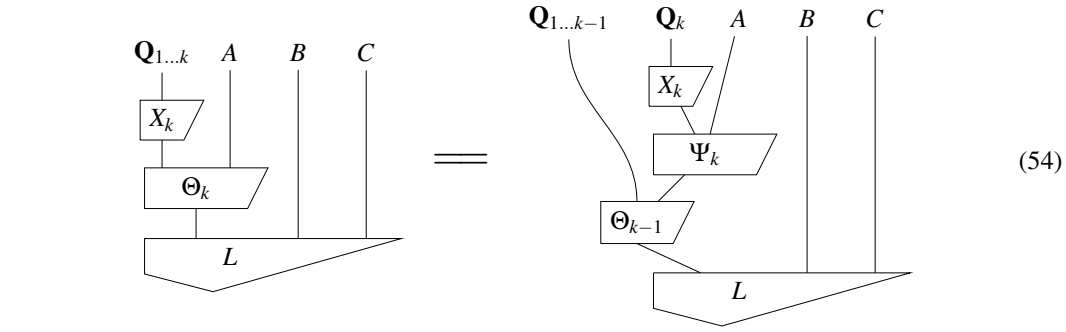
and the same holds with (X'_ℓ, X'_k) replaced by (X'_ℓ, Z'_k), (Z'_ℓ, X'_k), or (Z'_ℓ, Z'_k). Also, as noted at the beginning of section C.3, the gates that define $\Psi_{A,k}$ (see diagram (19) can each be pushed through L with error term $N\sqrt{\varepsilon}$. Therefore by Lemma 13,



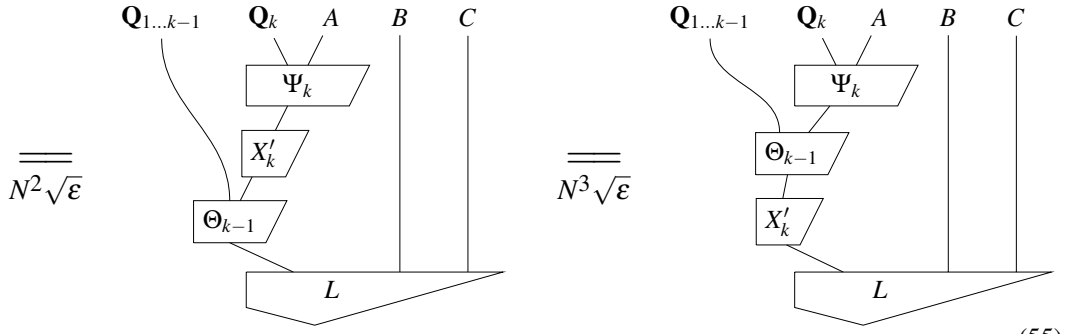
(53)

We therefore have the following, in which we first apply Proposition 8 with Lemma 7, and then apply

Lemma 13.



(54)



(55)

An analogous statement holds with X replaced by Z. Applying the isometries $\Psi_{k+1}, \ldots, \Psi_N$ in order now completes the proof.

Breiner, Spencer; Kalev, Amir; Miller, Carl.
"Three-Party Quantum Self-Testing with a Proof by Diagrams."
Paper presented at 15th International Conference on Quantum Physics and Logic (QPL 2018), Halifax, Canada. June 3, 2018 - June 7, 2018.

White Rabbit-Based Time Distribution at NIST

J. Savory, J. Sherman, and S. Romisch
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80305
joshua.savory@nist.gov

***Abstract*— The National Institute of Standards and Technology (NIST) produces a real-time realization of UTC(NIST) which is used to contribute to Coordinated Universal Time (UTC) and as a source for accurate time in the USA. The atomic clocks contributing to the time scale ensemble, the time transfer systems used to contribute to UTC and the distribution system used to disseminate UTC(NIST) to remote users are located in different parts of the NIST campus, far from each other and from the UTC(NIST) reference point. Since the physical inputs to these systems are not collocated within the campus, an accurate and stable infrastructure for time signal distribution is required. Currently, the local delays need to be known with an uncertainty of a few hundreds of picoseconds to avoid compromising the ultimate accuracy of the time transfer link's calibrations. Previously, coaxial cables or a commercial fiber-based frequency transfer system implemented by amplitude-modulation of a laser source were used to distribute signals between on-site locations, and clock trip calibrations were performed to measure the delays experienced by these signals [1]. The capability of WR-based time transfer systems to provide an on-time, accurate remote copy of its input pulse-per-second (PPS) signal made it a very appealing alternative to our previously implemented distribution system, which required time consuming re-calibration following instances of temporary signal interruptions. In this paper, we evaluate the use of WR-based time and frequency transfer within the NIST campus and verify its calibration procedure using a clock trip protocol [1].**

Keywords—Time Transfer, White Rabbit

I. Introduction

White Rabbit (WR) technology was initially developed at CERN, the European Organization for Nuclear Research, to produce a delay-compensated timing signal at remote locations for accelerator control and timing [2]. Information about the WR technology used at CERN is open source [3] and the hardware has been developed commercially. The base specifications of a calibrated WR time link are sub-nanosecond accuracy with picosecond stability, compatibility with commercial network hardware at fiber lengths of up to 10 km, and support for up to 1,000 nodes. The active compensation for the length and perturbations in the fiber is done using two-way time transfer through a combination of Precision Time Protocol (PTP) synchronization, Synchronous Ethernet (SyncE) synchronization and digital dual-mixer time difference phase detection.

At NIST, the time transfer systems used to contribute to UTC and the distribution system used to disseminate UTC(NIST) to remote users are located in different parts of the NIST campus, far from each other and from the UTC(NIST) reference point. Pulse distribution over long distances through coaxial cable is prohibitive due to attenuation, environmental, and impedance issues [4]. In comparison with other time transfer technologies, WR links offer the additional advantage that they can be calibrated to remove the propagation delays and produce an on-time remote secondary reference plane that is synchronized with the primary reference plane. While less stable than high-performance fiber-based frequency transfer systems over short averaging intervals, WR links maintain phase continuity over disruptions in the primary reference signal, and exhibit better long-term time stability. To determine the feasibility of using WR-based time transfer at NIST, its frequency and time stability were evaluated using a loopback measurement between two buildings on campus; environmental sensitivity of the master and remote transceivers were evaluated using an environmental chamber; and the calibration accuracy was evaluated by comparing it to several clock trip calibrations.

II. NIST's White Rabbit Fiber Link Layout

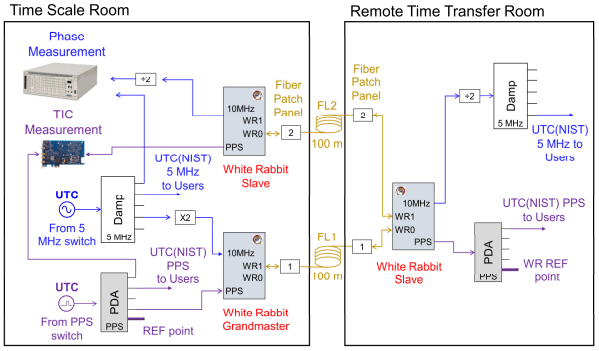


Fig. 1. Layout of the WR transfer system used to distribute time and frequency from the NIST timescale room to the remote time transfer room. The following items are abbreviated in the figure: Pulse Distribution Amplifier (PDA), Distribution Amplifier (Damp), and Coordinated Universal Time (UTC).

The WR link deployed at NIST uses the WR-LEN modules produced by Seven Solutions*. For the purposes of this work, a WR time link consists of a grandmaster module, which is sourced with a local 10 MHz and a pulse-per-second (PPS) reference, a slave module which outputs a 10 MHz and a PPS signal at a remote location and a single (bi-directional) fiber connecting the two modules. Each time module contains two laser ports: a module can be configured to act as a slave on one port and master on another allowing modules to be daisy

Savory, Joshua; Sherman, Jeffrey; Romisch, Stefania.
"White Rabbit-Based Time Distribution at NIST."
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

chained. The network hardware implemented in the NIST link has the grandmaster and master modules transmitting at a wavelength of 1310 nm, and the slave modules transmit at 1550 nm.

The layout of the WR fiber link at NIST is shown in Fig. 1. The link runs approximately 100 meters between the NIST time scale room and time transfer room where the time transfer equipment is housed. A WR grandmaster module, located in the time scale room, is sourced with a 10 MHz and a PPS signal directly from the UTC(NIST) signal plane. The grandmaster module is connected by Fiber Link 1 (FL1) to a remote slave module in the time transfer room, which reproduces the UTC(NIST) PPS and 10 MHz signals. These signals are used to source a TWSTFT (Two-Way Satellite Time and Frequency Transfer) Earth station and GPS (Global Positioning System) receiver for dissemination of UTC(NIST). The second fiber port on the remote WR module is run in master mode and is used to connect back to a local slave module in the time scale room via Fiber Link 2 (FL2). The outputs of this module are used for monitoring and for performing stability measurements of the WR link.

III. Time Transfer System Comparison

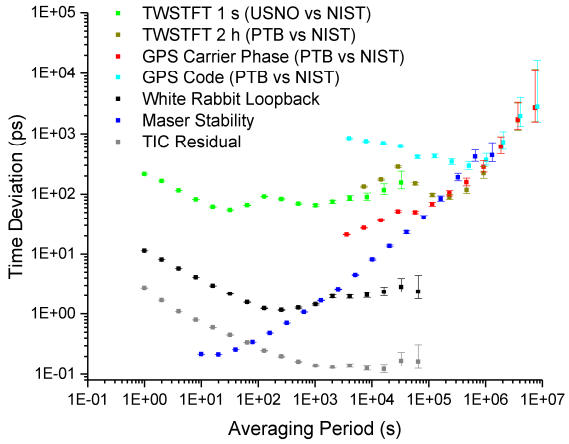


Fig. 2. Comparison of the time deviation (TDEV) as function of averaging period between several time transfer systems at NIST. The WR loopback PPS measurement (see Fig. 1) is shown in black and the residual of the time interval counter used to perform this measurement is shown in gray. Also shown are a TWSTFT comparision of UTC(NIST) and UTC(USNO) at a time interval of 1 s (green), a TWSTFT comparision of UTC(NIST) vs UTC(PTB) at a time interval of 2 h (dark green), a comparion of UTC(NIST) vs UTC(PTB) via GPS code (cyan) and carrier phase (red), and a typical maser stability (blue).

The stability of the transmitted PPS was evaluated over the course of several days by comparing the UTC(NIST) PPS and the returning WR link's PPS on a GuideTech (GT668PCIe-1)* time interval counter as shown in Fig. 1. A comparison of the of this measurement to typical results from NIST's GPS and TWSTFT links can be seen in Fig. 2. The stability of the GPS and TWSTFT links are measured through time scale comparisons, and therefore also contain the noise of the scales themselves. The transfer system noise will dominate in the short-term and the scale-to-scale noise determines the long-term performance of the comparison. A typical maser stability is also included in this plot to show approximately at what averaging period the clock noise dominates, thus allowing the clock signal to be transferred without corruption. The WR link PPS's stability is below 20 ps at all averaging periods, and is below the clock noise for averaging periods greater than 1000 s. Furthermore, as the short-term noise of the WR link is white phase it is well below the minimum time deviation (TDEV) of GPS carrier phase (~20 ps) for averaging periods greater than 10 s.

IV. Long-Term Frequency Stability Evaluation

The WR link's stability is continually monitored by dividing the 10 MHz signal from the local WR slave module to 5 MHz and comparing it with the UTC(NIST) 5 MHz signal (see Fig. 1) on a multichannel dual-mixer phase measurement system from Microchip (TSC 12030)*. A measurement of the phase difference for the loopback over ~100 days is shown in Fig. 3 and a phase drift of approximately 40 ps is observed. This corresponds to a long-term frequency offset of ~5×10^{-18} and sets the frequency accuracy of the link. The observed phase drift appears to be decreasing in magnitude in the last 20 days of data and could be due to aging in any one of the components (isolation amplifiers, frequency divider, frequency doubler, WR hardware, etc.) involved in the measurement chain. With this performance, in order to ensure that the secondary reference point's time offset is within 200 ps of UTC(NIST), the WR link should be recalibrated at least once a year.

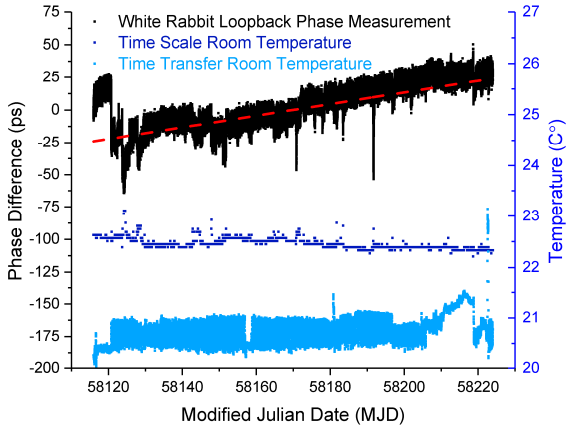


Fig. 3. Plot of the long-term phase difference (left axis) over time of the White Rabbit 10 MHz loopback is shown in black and the temperatures (right axis) of the time scale and time transfer rooms are shown in dark blue and light blue respectively.

The modified Allan deviation (MDEV) of the loopback phase data is shown in Fig. 4 together with that of a typical maser signal at a measurement bandwidth of 0.1 Hz (10 s). After 500 s of averaging, a maser signal would be observable on the 10 MHz WR link as the link noise at this averaging interval is below the noise of the maser. The link stability is ~1×10^{-16} after 100,000 s of averaging and has not yet reached the flicker floor. However, the overall frequency accuracy of the link will be set at ~5×10^{-18} by the phase drift observed in Fig. 3.

In Fig. 3, a correlation is observed between phase excursions in the loopback measurement and fluctuations in temperature of

Savory, Joshua; Sherman, Jeffrey; Romisch, Stefania.
"White Rabbit-Based Time Distribution at NIST."
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

the room which houses the WR modules. The WR module's case is designed to be the heat sink for the module itself, hence the observed temperature dependence. In the next section, the environmental sensitivity of the time modules is explored in a more controlled setting.

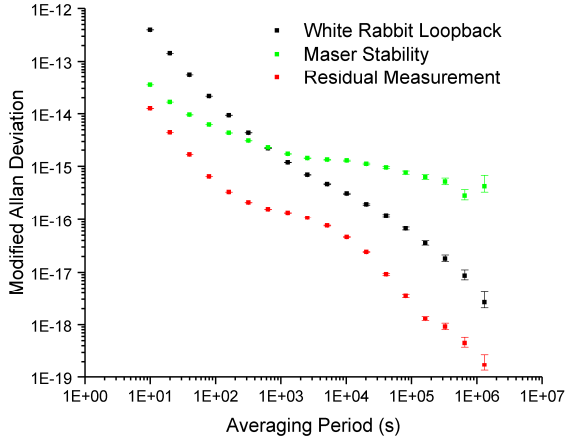


Fig. 4. Comparison of the modified Allan deviation (MDEV) as a function of averaging period for the WR link 10 MHz loopback (black) measurement, a typical maser stability (green) and a residual measurement (red) performed on the phase measurement system used to collect both sets of phase data.

V. Environmental Sensitivity Measurements

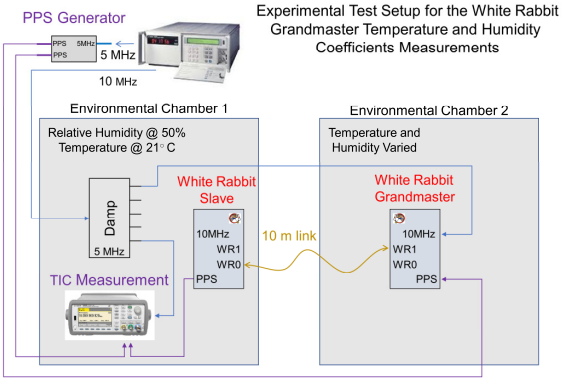


Fig. 5. Test setup used to evaluate the WR grandmaster modules' environmental sensitivity.

To quantify the WR time modules' environmental sensitivities, two modules were placed in separate environmental chambers. The experimental setup used to evaluate the grandmaster module is shown in Fig. 5. The grandmaster module was sourced with synchronous PPS and 10 MHz signals from a 5071A Cesium clock from Microchip* and placed in the test environmental chamber whose temperature and humidity were varied. The slave module was connected to the grandmaster module using a 10 m fiber link and placed in the control chamber which remained static in temperature and humidity for the duration of the test. The phase change in the WR slave module PPS against the source PPS was measured on a time interval counter from Spectracom (Pendulum CNT-91)*. To determine the temperature coefficient, the test chamber was varied by 2ºC steps (21ºC to 23ºC and back). Similarly, to measure the humidity coefficient the humidity in the test chamber was stepped in increments of 10% RH (50% RH to 60% RH and back). In both cases, the chamber was left at each temperature and humidity setting for a time interval of at least 1 day. A similar test setup was used to evaluate the environmental sensitivity of the WR slave module in which the slave module was placed in the test chamber and the grandmaster in the control chamber.

The results of these measurements are summarized in Table 1. No significant humidity coefficient was observed for either the grandmaster or slave module. However, a large temperature coefficient of 29 ps/ºC was observed for the grandmaster module, which is approximately seven times larger than the one for the slave module and that of a typical isolation amplifier (5 ps/ºC). This result coincides which the temperature correlation observed in Fig 3. which shows a sensitivity bias towards the time scale room's temperature.

Module Enviromental Sensitivity

Module	***Temperature Coefficient***	***Humidity Coefficient***
Grandmaster	29.0(2) ps/°C	0.41(3) ps/%
Slave	4.0(2) ps/°C	0.04(2) ps/%

Table 1. Table of measured WR time module phase sensitivity to changes in temperature (P_t) and humidty (P_h).

VI. Time Offset Calibration

The WR link must not only be stable but accurate as well. The WR modules allow for the alignment in time of the PPS at the primary and secondary reference points. To accomplish this the WR time modules were initially calibrated according to the procedure in [5] to remove time offsets in the WR link due to fiber length and internal electrical delays. To account for external delays in the PPS distribution chain and align the primary and secondary reference points, clock trip [6, 7, 8] and WR trip measurements were performed to determine the remaining PPS delays. The value from these measurements were programmed into WR grandmaster and slave module to do the final alignment.

In a clock trip, a traveling clock is moved between the primary and secondary reference point, and the time difference between the reference point and the traveling clock is measured at each location. Using prior information about the traveling clock's frequency and stability, the time offset between the two reference points and the associated uncertainty can be calculated. By comparing the time difference between the traveling clock and the primary reference point measured at the beginning and end of each clock trip, the measurement's systematics can be accessed in what is called the closure measurement. A successful closure measurement should be statistically consistent with zero. The full details of this procedure and measurement parameters used in this work are in [1].

Similarly, a WR calibration trip consists of using a grandmaster and slave pair of WR time modules to measure a secondary point against a primary reference point. In this procedure, the grandmaster WR time module is sourced by a PPS and a 10

Savory, Joshua; Sherman, Jeffrey; Romisch, Stefania.
"White Rabbit-Based Time Distribution at NIST."
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

MHz synchronous with the primary reference signals at the primary reference point's location where it remains for the duration of the measurement. Initially, the slave module is connected to the grandmaster using a local fiber (fiber A) and the initial time difference ($\Delta\theta_I$) between the slave's PPS output and the primary reference point's PPS are measured on a time interval counter. The slave module is then moved to the location of the secondary reference point connected to the grandmaster using a second calibration fiber link (fiber B) and the mid-point time difference ($\Delta\theta_M$) between the slave's PPS output and the secondary reference point's PPS is measured at the remote location. For the closure measurement, the slave module is then returned to the location of the primary reference point and the final time difference ($\Delta\theta_F$) between the slave's PPS output and the primary reference point's PPS is measured. The time offset (D) of the secondary reference point is calculated from these measurements

$$D = \Delta\theta_M - \frac{\Delta\theta_I + \Delta\theta_F}{2}. \quad (1)$$

The closure (C) of the WR trip can be expressed as

$$C = \Delta\theta_F - \Delta\theta_I \quad (2)$$

and the uncertainty is simply the propagated error of the time difference measurements. The WR trip is simpler than the clock trip because no time prediction is required as the WR link time signals are synchronous with the primary reference signals.

The time difference measurements for the WR trip were all performed with a time interval counter from Spectracom (Pendulum CNT-91)*. Each time interval measurement was averaged over a period of 100 s which corresponds to a statistical uncertainty of ~4 ps at this average period (see Fig. 2). At each location, the link was re-established and the modules' temperature allowed to settle for at least 30 minutes before performing the time difference measurements. After the initial calibration, the WR link's time delay is to first order independent of the fiber link length. To further reduce this error the lengths of the two calibrations fibers (A and B) were chosen to be approximately equal.

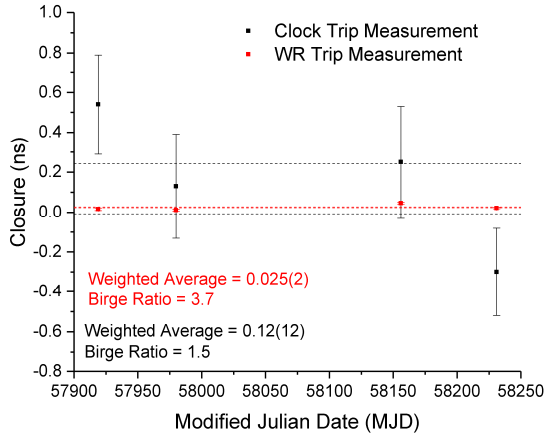


Fig. 6. Result of the closure measurements at the MJD they were performed for clock trip (black) and WR trip (red) protocols.

The following WR trip and clock trip measurements were preformed to validate a successful recalibration of the WR link and to evaluate the calibration methods. The measurements were performed at different dates and should be viewed as independent and not as a reflection of the link's stability. A comparison of the closure measurements obtained for the clock trip and WR trip calibrations are show in Fig. 6. As previously observed in [1], the clock trip's closure measurements are statistically consistent with zero and have a Birge ratio consistent with ~1 based on a small number of measurements [9]. On the other hand, the WR trip measurements have a significant non-statistical bias and non-statistical scatter (Birge ratio >> 1). To account for this the WR time offset uncertainties are scaled by the Birge ratio (3.7) and summed with half of the associated closure measurement. The most likely contributor to the observed bias and scatter is the large temperature coefficient of the grandmaster module. During the calibration process, it is difficult to maintain a constant case temperature as the equipment is moved between different locations and the link re-established.

For both methods, the time offset measurements of the secondary reference point in the time transfer room are shown in Fig. 7. At all measurement dates and with both methods the measured time offset was less than 200 ps and in most cases statistically consistent with the zero. Despite being scaled based on the closure measurements, the uncertainties for the WR trips are still a factor of ~5 smaller than the clock trip uncertainties. The difference between the time offsets measured with the two protocols is show in Fig. 8. No statistically significant bias between the two methods is observed at the one sigma level. Based on a conservative interpretation of these results, the WR trip protocol can be used to calibrate the WR link at sub-100 ps accuracies with a single measurement. If the systematics of a WR trip can be better constrained, time offset measurements at the 10 ps level should be possible.

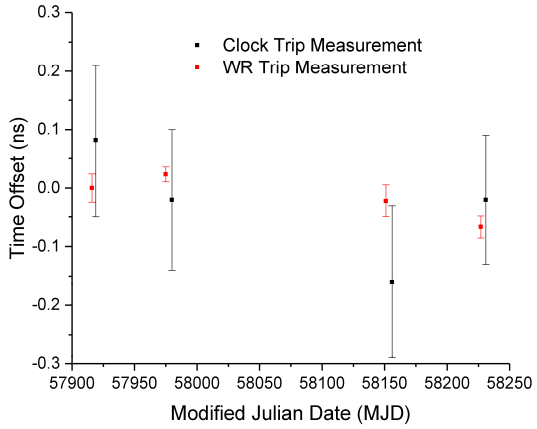


Fig. 7. Result of the time offset measurements of the time transfer room reference point at the MJD they were performed for clock trip (black) and WR trip (red) protocols.

Savory, Joshua; Sherman, Jeffrey; Romisch, Stefania.
"White Rabbit-Based Time Distribution at NIST."
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

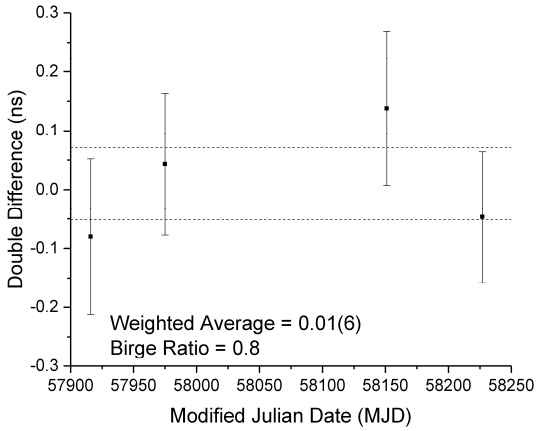


Fig. 8. Time offset difference obtained with the two different calibration protocols (WR trip – clock trip) at the MJD they were performed for the time transfer room reference point.

VII. Conclusions

The WR hardware has met the accuracy and stability requirements to provide UTC(NIST) to the time transfer systems used to contribute to UTC that are in different locations on the NIST campus. Measurements performed on the PPS WR loopback show that WR link's time stability is below 20 ps at all averaging periods, and thus stable enough to distribute UTC(NIST) throughout the NIST campus. Time offset measurements done with the clock and WR trips calibration protocol are in high agreement and demonstrate that WR link can be calibrated and measured to an accuracy of < 100 ps. Furthermore, a long-term measurement of the 10 MHz WR loopback shows a frequency accuracy of ~5×10^{-18} which means that WR link need only be recalibrated once a year. The WR trip protocol has also been investigated and shown to offer an alternative method to calibrate remote reference points with a factor of five smaller uncertainties than that of a clock trip. The accuracy and stability of the WR modules could be further improved by reducing the temperature coefficient of the grandmaster module which is currently under evaluation [10].

Disclaimer

This paper includes contributions for the U.S. Government and it is not subject to copyright.

* Certain commercial equipment, instruments, or materials are identified in this paper for informational purposes only. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Acknowledgments

The authors would like to thank Tom Parker, Bijunath Patla, Jian Yao, and Victor Zhang for providing data on the NIST time transfer systems.

References

[1] J. Savory, S. Romisch, L. Hernandez, and K. Maurer, "Local Distribution and Calibration of Timing Signals at NIST," Proceedings of the 48th Annual Precise Time and Time Interval Systems and Applications Meeting, Monterey, California, pp. 326–342, January 2017.

[2] J. Serrano, P. Alvarez, M. Cattin, E. Garcia Cota, J. Lewis, P. Moreira, T. Wlostowski, G. Gaderer, P. Loschmidt, J. Dedič, R. Bär, T. Fleck, M. Kreider, C. Prados, and S. Rauch, "THE WHITE RABBIT PROJECT", Proceedings of the 12th International Conference On Accelerator And Large Experimental Physics Control Systems, Kobe, Japan, pp. 93–95, October 2009.

[3] White Rabbit Open Hardware Repository: Technical Presenations https://www.ohwr.org/projects/white-rabbit.

[4] M. Siccardi, D. Rovera, and S. Romisch, "Delay measurements of PPS signals in timing systems", *Frequency Control Symposium (IFCS)*, 2016 IEEE International, New Orleans, LA, 628, May 2016.

[5] G. Daniluk. (2014, Aug. 13). White Rabbit Calibration Procedure Version 1.0, CERN BE-CO-HT [Online]. Available: http://www.ohwr.org/documents/213.

[6] L. N. Bodily and R. C. Hyatt, "'Flying Clock' Comparisons Extended to East Europe, Africa and Australia", *Hewlett-Packard Journal*, 19, 12, 1967.

[7] H. Hellwig and A. E. Wainwright, "A portable rubidium clock for precision time transport," *Proceedings of the 7th Annual Precise Time and Time Interval Systems and Applications Meeting*, Washington, D.C., 143, December 1975.

[8] L. G Rueger and R. A. Nelson, "Portable Hydrogen maser clock time transfer at the subnanosecond level," *Proceedings of the 19th Annual Precise Time and Time Interval Systems and Applications Meeting*, December 1987.

[9] T. Parker "Update on a Comparison of Cesium Fountain Primary Frequency Standards" *2011 Joint Conference of the IEEE International Frequency Control Symposium and European Frequency and Time Forum*, San Francisco, July 2011.

[10] White Rabbit Open Hardware Repository: WRS Low Jitter Daugherboard https://www.ohwr.org/projects/wrs-low-jitter/wiki.

Savory, Joshua; Sherman, Jeffrey; Romisch, Stefania.
"White Rabbit-Based Time Distribution at NIST."
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

H. Roßnagel et al. (Eds.) (Hrsg.): Open Identity Summit 2019,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 1

Smart Contract Federated Identity Management without Third Party Authentication Services

Peter Mell[1], Jim Dray[2] and James Shook[3]

Abstract: Federated identity management enables users to access multiple systems using a single login credential. However, to achieve this a complex privacy compromising authentication has to occur between the user, relying party (RP) (e.g., a business), and a credential service provider (CSP) that performs the authentication. In this work, we use a smart contract on a blockchain to enable an architecture where authentication no longer involves the CSP. Authentication is performed solely through user to RP communications (eliminating fees and enhancing privacy). No third party needs to be contacted, not even the smart contract. No public key infrastructure (PKI) needs to be maintained. And no revocation lists need to be checked. In contrast to competing smart contract approaches, ours is hierarchically managed (like a PKI) enabling better validation of attribute providers and making it more useful for large entities to provide identity services for their constituents (e.g., a government) while still enabling users to maintain a level of self-sovereignty.

Keywords: federated identity management; authentication; blockchain; smart contract

1 Introduction

Federated identity management (FIM) enables users to access multiple systems using a single login credential. In industry implementations (e.g., with Amazon, Google, and Facebook authentication[4]), multiple entities collaborate such that one entity in the collaboration can authenticate users for other entities; it requires complex interactions to enable a user to perform a business interaction with some 'relying party' (RP) (e.g., a business) and have the authentication performed by a 'credential service provider' (CSP) (the entity performing the authorizations) [TA18]. It may involve redirecting a user from an RP to a CSP and then back to the RP post-authentication with the CSP communicating with both the user and RP. CSPs likely will charge for this service while being able to violate the privacy of users by seeing with which RPs they interact. Complicating matters further, FIM often supports the transferring of user attributes (e.g., age) to an RP to support a business interaction.

[1] National Institute of Standards and Technology, Computer Security Division, 100 Bureau Drive Gaithersburg, MD 20899 U.S.A. peter.mell@nist.gov
[2] National Institute of Standards and Technology, Computer Security Division, 100 Bureau Drive Gaithersburg, MD 20899 U.S.A. james.dray@nist.gov
[3] National Institute of Standards and Technology, Computer Security Division, 100 Bureau Drive Gaithersburg, MD 20899 U.S.A. james.shook@nist.gov
[4] Any mention of commercial products is for information only; it does not imply recommendation or endorsement.

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

In this work, we provide an identity management system (IDMS) that provides FIM such that a user can authenticate and transfer attributes to an RP without the involvement of a CSP (thereby heightening privacy and reducing costs). We accomplish this through leveraging a smart contract running on a blockchain[5]. User to RP interactions do not need to transact with the smart contract, they simply use data from a copy of the blockchain. Thus, there is no need for the user or RP to wait for blockchain blocks to be published or to pay blockchain transaction fees. User to RP communications are extremely fast and free.

Our IDMS is hierarchically managed enabling authorities to manage user accounts and associate attributes with accounts. However, users are granted a degree of self-sovereignty; a user must approve added attributes and can view and delete their data. Privacy is maintained by either adding only hashes of attributes to user records, by only adding data encrypted with the user's public key, or by only adding references to external and secured databases that house user attribute data. We emphasize that user to RP interactions are completely private, something not possible in current systems using a CSP for authentication.

We implemented our IDMS on the Ethereum platform [Eth]. Charges are only incurred when creating and updating user accounts, which is something that is relatively rare compared to a user freely and regularly interacting with RPs. Also, user account update functions are very cheap, all costing less than $0.09 USD (as of September, 2018). We note that other FIM smart contract systems are in development, but ours differs primarily in being a managed approach that still provides a degree of user self-sovereignty. This provides advantages in having authoritative identity attributes for users and having the ability to validate attribute providers.

The rest of the paper is structured as follows. Section 2 describes the overall contract design and section 3 describes the attribute field design. Then section 4 outlines the core functions of the IDMS system: authenticating users and passing attributes. Section 5 provides an example, section 6 discusses our implementation, section 7 explains why we use smart contracts, and section 8 enumerate achieved security properties. Section 9 provides the related work and section 10 our conclusions.

2 IDMS Contract Design

Our IDMS is implemented within a smart contract accessed by five types of entities: the IDMS owner, account managers, attribute managers, users, and RPs (shown in figure 1). The first four issue transactions to the blockchain to manage user accounts (relatively rare events). Users and RPs use public blockchain data to authenticate a user and pass attributes (the more common events). Both the managers and users have IDMS accounts. Manager data is publicly readable while user data is kept private using hashes and encryption.

[5] See [Yag+18] for an overview of blockchain and smart contract technology.

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

Smart Contract: The smart contract is modeled as being immutable; once deployed, it is not owned and is its own entity. Alternately, it may be coded for the IDMS owner to update it with participant agreement (e.g., a voting mechanism) or after a notification period (allowing participants time to withdraw from the IDMS if they disapprove of the changes).

IDMS Owner: The IDMS owner is limited by the contract to authorize and deauthorize managers. For authorization, an entity creates a blockchain account, gives their public key to the owner, and the owner directs the contract to create an IDMS manager account for that public key. For deauthorization, the account record is marked as invalid. For each created manager, the owner specifies one or more descriptor fields. This should follow a standard nomenclature to enable automated evaluation of these fields by other entities (e.g., by RPs).

Account managers: Account managers authorize user accounts in an analogous manner as the IDMS owner does for managers. User records are pseudonymous, they contain no identifying information. An account manager can only perform deauthorization on accounts they created. If a user's private key is lost or stolen, the account manager may authorize a new account for the user using a new public key generated by the user and deauthorize the old account. The IDMS owner can require the account managers to perform identity proofing at some level, confirming that users are whom they claim to be. The contract can require a subset of the collected attributes to be posted to the user account. We refer to such attributes as 'identity attributes'; they can be updated at any time by the account manager.

Attribute managers: Attribute managers add attributes to users' accounts. However, users must first grant them permission. They may revoke any attributes previously added.

Users: Users may unilaterally delete non-identity attributes (to avoid them changing their identity). They may also delete their IDMS account completely. As mentioned previously, they must authorize any attribute manager to add attributes to their account.

RPs: RPs keep a local copy of the contract state, extracted from the blockchain, and execute contract 'view' functions on that copy to enable reading the contract data. They do not have accounts on the contract or transact with the contract.

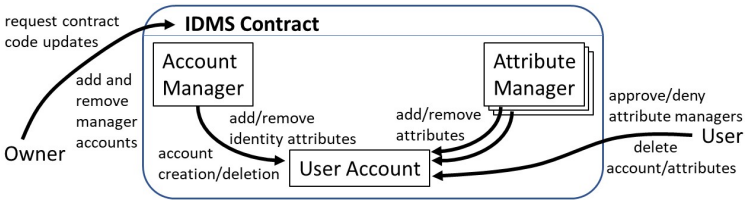


Fig. 1: IDMS Contract Design Relative to a Single User

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

3 IDMS Attribute Field Design

An important design element is the attribute field. Each field has a hash of a user attribute (put there by the applicable account manager or an attribute manager). If the actual attribute data is included to allow for easy user retrieval (which is not necessary), it is encrypted with a secret key that is then encrypted with the user's public key to preserve user privacy. It is expensive to store data on a blockchain; if the data is large (e.g. video or image files), an off-blockchain location of the data may be posted to the attribute field. This might be used, for example, with images of physical credentials such as driver's licenses, visas, social security cards, and passports. Note that the source of each attribute field is public to allow RPs to check the authority behind each user provided attribute.

Field Name	**Field Description**
ManagerPublicKey	Public key of manager that posted the attribute
Identity	Boolean to indicate if this is an identity attribute
EncryptedSecretKey	Secret key encrypted with the user's public key
Descriptor	Encrypted description attribute data
Data	Encrypted attribute data
Location	Location for downloading data
Hash	Hash of the unencrypted descriptor and data

Tab. 1: Contents of an Attribute Field

To accomplish this, we use the attribute field structure shown in table 1. The 'ManagerPublicKey' field is the public key of the manager that posted the attribute to the user's account. This key can enable anyone to look up the manager in the IDMS using the publicly available blockchain data. Manager accounts contain only unencrypted attributes so that anyone can verify who posted an attribute. Note that only the contract owner can authorize a manager and populate its data fields, thus the unencrypted attributes within a manager's account are considered authoritative. The 'Identity' field is a boolean indicating whether or not an attribute is an identity attribute. The 'EncryptedSecretKey' is the secret key that was used to encrypt the attribute descriptor and data fields. The 'Descriptor' field is an encrypted field that explains what the attribute data field contains[6]. The optional 'Data' field contains encrypted attribute data (these must be appended with a nonce prior to encryption to prevent guessing attacks when the attribute space is limited). The optional 'Location' field identifies a public location where the encrypted attribute data is available. The 'Hash' field is a hash of the unencrypted Data field appended with the unencrypted Descriptor field. This enables an RP to verify that a user is providing them the correct data and descriptor fields for a particular source. Note that if neither the Data or Location fields are provided, the user must maintain copies of the data for which the relevant hashes are posted.

[6] Implementations of this should standardize on a set of descriptors and a format for the data field to promote automated processing of the attribute data.

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

4 IDMS Core Functions

In this section we will describe the core functions for our conceptual IDMS system: 1) authentication of users and 2) secure transmission of user attributes. A key design feature is that the user and RP can achieve this without any interaction with a third party (they don't even need to transact with the smart contract). However, the user needs access to their attribute descriptors and data. These could be maintained by the user, downloaded from the blockchain (if stored in encrypted form in the user's record), or downloaded and decrypted from the location specified in the location field of the user's record. The user will also need to maintain their private key. This could be done in a hardware dongle to promote security and portability between devices, but could also be copied to multiple devices if desired.

The RP will need access to a copy of the blockchain on which the contract is being executed (which is publicly available through the blockchain peer-to-peer network). They need only store the small portion relevant to the contract data. This must be a version recent enough as to have a hash of the attributes that the user will provide to the RP. Note that the RP does not need a blockchain account and the user will not need to transact with their blockchain account for these core functions (they do so only to maintain their contract user record).

4.1 IDMS Authentication

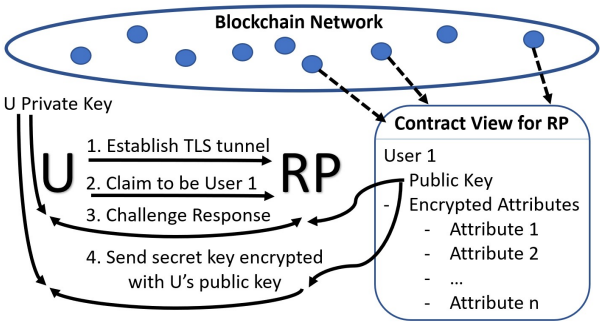


Fig. 2: Example User to RP Authentication Function

Our first core function enables U to authenticate to some RP_1 given that RP_1 can access U's public key from the IDMS data on the public blockchain. This could be done through many approaches; here we present a method using Transport Layer Security (TLS). Our approach is similar to using TLS with client-side certificates, except that in our scenario no such client-side certificate exists. We achieve this by creating a TLS session, but within that session adding an additional challenge response mechanism followed by RP_1 generating a final symmetric key used for a second encrypted tunnel within the original TLS tunnel.

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

With additional engineering, this tunnel within a tunnel approach could be replaced with the second 'challenge response' tunnel replacing the first TLS tunnel.

More specifically for our example approach, U establishes a TLS tunnel with RP_1. U then sends a message to RP_1 claiming to own account 'User 1' in the IDMS. RP_1 then accesses the IDMS account 'User 1' using its local copy of the blockchain and retrieves the posted public key. RP_1 sends a random challenge to U encrypted with the public key posted on the IDMS account. U decrypts this with his private key and sends the result to RP. If the correct value was returned by U, then U has proved ownership of account 'User 1'. Next, RP_1 encrypts a symmetric key with U's public key to use for the second encrypted tunnel and sends it to U. U obtains the symmetric key by decrypting with his private key. At this point both U and RP_1 have mutually authenticated and have established an encrypted tunnel. This process is shown in figure 2.

Note that in TLS, U produces the symmetric key used for the encrypted tunnel. However, in our secondary tunnel it is necessary that RP_1 produce the symmetric key and encrypt it with U's public key to avoid a man-in-the-middle attack. We must prevent RP_1 from being able to masquerade as U while accessing some RP_2 (because RP_1 could answer RP_2's challenge using a response obtained by issuing the same challenge to U).

4.2 IDMS Attribute Transfer

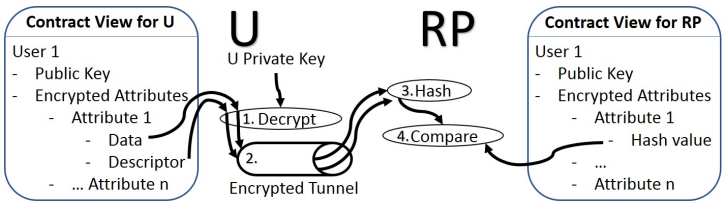


Fig. 3: User to RP Attribute Transfer Function

Our second core function enables U to send attributes to RP (e.g., personal information necessary to complete some interaction). U obtains a decrypted copy of an attribute descriptor and data (from a local store, from an encrypted version stored in the user's IDMS record, or from a server whose location is specified in the user's IDMS record). U sends the descriptor and data to RP. RP hashes a concatenation of the data and descriptor and then verifies that the result matches a hash on the user's IDMS record. The RP can then use the 'ManagerPublicKey' field in the matching attribute record to evaluate the attribute source.

The manager accounts have unencrypted descriptor fields populated by the owner to enable an RP to automatically evaluate the authority of a manager account (e.g., that the manager issuing a drivers license really is the correct government agency). By the owner populating

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

these public manager descriptor fields with a standard nomenclature, automated evaluation by RPs of a manager's authority can be enabled.

5 Example Use Case

A government deploys an instance of our IDMS contract to a blockchain and is the owner. The owner authorizes account manager entities to perform identity proofing and add users. This is likely organizations already performing related activities, such as banks and local governments. A user Bob goes to his bank to have an account created in the IDMS. After providing the necessary documentation, he is granted an account. The owner also authorizes a university as an attribute manager with the descriptor fields 'university' and 'University of Corellia'. The former is a standardized descriptor to enable automated processing while the latter provides the name of the specific university (note that how to create ontologies of descriptors is out of scope of this work). Bob then requests that the University of Corellia post his degree to his IDMS account. Bob must first prove to the university using the core IDMS functions that 1) he owns the account and 2) that the account is for his identity by passing them identity attributes. Bob then transacts with the IDMS contract to give the university permission to post attributes to his account. The university gives Bob a digital image of his degree and also posts an attribute on Bob's IDMS account with a hash of the digital image and a location field indicating where Bob can login and download the image off of university servers (in case Bob loses the originally provided digital image). The university posts a second attribute indicating his grade point average (GPA). Since this is a small data field, it is encrypted along with a standard sized nonce and placed inside the attribute field. Bob can download this anytime off of the blockchain and use his private key to decrypt it. Bob then applies for a job with Ally, who wants proof that Bob graduated with a minimum GPA. Bob uses the core IDMS functions to prove that 1) he owns the IDMS account and 2) that the account contains the attributes necessary to convince Ally that Bob received a degree and graduated with a sufficient GPA. When Ally receives and verifies the attributes sent by Bob, she then checks the descriptor fields associated with the attributes. She verifies that the attributes were provided from a university using the first descriptor field and she reads off the specific university using the second descriptor field.

6 Implementation Details and Empirical Study

We implemented our IDMS using a smart contract running on the Ethereum platform [Eth] and created apps to interact with the smart contract. The contract implements all of the functionality described in section 2 and it contains methods to support the core functions described in section 4. Note that we left for future work the implementation of the off blockchain U to RP interactions.

We tested all contract interactions described in sections 2 and 4. There were two types of interactions: transactions and views. Transactions are function calls that change the state of

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

the contract; they thus must be submitted to the miners so that the changes can be stored on the blockchain. Views are function calls that look like transactions except that they do not alter the state of the contract; they thus can be executed locally by a node that has a copy of the blockchain. This makes their use free and fast. Table 2 lists the implemented functions.

Function	**Type**	**Permitted Role**	**Gas**	**Ether**	**USD**
Add Manager	Transaction	Contract Owner	66632	2.0E-4	$0.03
Delete Manager	Transaction	Contract Owner	17677	5.3E-5	$0.01
Add User Account	Transaction	Account Manager	94562	2.8E-4	$0.05
Delete User Account	Transaction	Account Manager	65020	2.0E-4	$0.03
Add Attribute	Transaction	Managers / Users	182045	5.5E-4	$0.09
Delete Attribute	Transaction	Managers / Users	33017	9.9E-5	$0.02
Permit Attribute Manager	Transaction	Users	45151	1.4E-4	$0.02
Deny Attribute Manager	Transaction	Users	15283	4.6E-5	$0.01
Compare Hash	View	Public	0	0	$0
View Attribute	View	Public	0	0	$0
View Public Key	View	Public	0	0	$0

Tab. 2: IDMS Contract Functions and Costs ($219.01 USD/Ether as of September 27, 2018)

Note that the view functions are used by users and RPs for their interactions. The transaction functions are only used to set up the IDMS data structures. Thus, normal operation of our IDMS is extremely fast and does not cost anything. Creating user accounts and updating them with attributes costs a modest amount of funds (e.g., less than $1 USD), but such activities are relatively rare compared to users interacting with RPs.

7 Reasons to use a Smart Contract

Use of the smart contract promotes trust in the system while providing a convenient vehicle for data distribution and update of a distributed and resilient data store. The smart contract code is publicly viewable and immutable, thus all participants know how it will operate and all entities are constrained to their roles. In particular, the owner is limited to just creation and deletion of manager roles; no access to user accounts is provided. The blockchain peer-to-peer network makes it convenient to distribute the IDMS data to participating entities. This also provides transparency and audit-ability for all IDMS transactions. Since the user to RP interactions don't modify the blockchain, this transparency doesn't cause a problem with user privacy. Lastly, the smart contract approach enables one to deploy an IDMS without the need to build and maintain any infrastructure.

8 Security Properties

We now summarize the security and privacy properties needed for our model and then explain how each security property is fulfilled by our IDMS and then discuss a residual weakness. The specific security properties are as follows:

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

1. User attribute data is encrypted such that only the user can decrypt it.
2. Users can securely share their attribute data with other parties.
3. Users can unilaterally remove their attributes.
4. Users can unilaterally remove their account.
5. Users can have multiple accounts in order to hide their association with certain attribute managers.
6. Account managers can only remove accounts that they created. Owners and attribute managers may not remove accounts.
7. Account managers can only modify the identity attributes for accounts they created.
8. Attribute managers may only place attributes if explicitly permitted by the relevant user.
9. Owners may only add and remove account/attribute managers and update the IDMS contract code.
10. IDMS contract code may only be updated by the owner following due process laid out in the contract (which is publicly available to all users of the contract).
11. Relying parties can trust account managers to perform identity proofing that binds real world entities to user accounts at a stated level of assurance.

These security properties are provided primarily by the contract itself. Except under conditions documented within the contract, the code is immutable. The code is also public so that users can verify that these properties will be held. The contract directly enforces security properties 3, 4, 6, 7, 8, 9, and 10. Key to this enforcement is for the smart contract to authenticate the party requesting a change. This is handled by the smart contract system, leveraging the accounts on the blockchain. Thus, our approach does not have to implement that part of the trust model.

Property 1 is enacted by the account and attribute managers when they place attributes on a user account. There is nothing in the contract to prevent the posting of unencrypted attributes, but there is no motive for a manager to do so and there could be repercussions (e.g., the owner could remove the manager from the IDMS).

Property 2 is enabled since our IDMS architecture provides a way for a user and RP to directly authenticate and pass attributes. All they need is to use a standard encrypted connection within which to execute our protocol.

Property 5 can be provided by a user's account manager. It is trivial to create additional accounts on blockchain systems, thus the user can do so easily. The account manager then simply creates an IDMS account with the public key associated with each of the user's accounts. Based on our empirical work, there may be a modest cost to create each account

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

(e.g., $0.05 USD). Also, we note that users are not required to pass RPs their identity attributes, enabling them to pass other attributes without revealing their identity. This can enable transactions to authenticate that a person has some attribute while staying anonymous. An example might be an online forum where only members of a certain organization can post messages but where the poster's identity is to remain anonymous.

Property 11 is achieved through the contract owner auditing the account managers to ensure that users are identity proofed at the required or advertised level of assurance. If account managers are non-compliant then the contract owner can revoke their accounts.

Despite these security protections, we note an important limitation. An account managers could use their knowledge of a user's identity attributes to create a clone identity for someone else. This is analogous to a government duplicating someone's passport but including a different picture to enable someone to act as someone else. To our knowledge this problem exists in the related schemes (discussed next) whenever attribute managers act maliciously.

9 Related Work

Many organizations are investigating using blockchain technology for identity management. Our approach is unique in providing a managed hierarchical approach with user self-sovereignty that can authoritatively validate attribute providers (or claim providers).

uPort: uPort is an 'open identity system for the decentralized web' [uPo18]. uPort users create and manage self-sovereign identities by creating Ethereum accounts linked to a self-sovereign wallet. Being unmanaged and fully self-sovereign, there is no entity identity proofing of user accounts [Lun+17]. Our approach differs in that it provides a managed solution that still provides a level of self-sovereignty. This managed aspect can enforce validation on the claim providers not possible in completely unmanaged systems.

SCPKI: The paper entitled 'SCPKI: A Smart Contract-Based PKI and Identity System' [AlB17] addresses the issue of rogue certificates issued by Certificate Authorities in traditional public key infrastructures. It proposes an alternative PKI approach that uses smart contracts to build a decentralized web-of-trust. The web-of-trust model is adopted from the Pretty Good Privacy (PGP) system [Gar95]. SCPKI supports self-sovereign identity by defining a smart contract that allows users to add, sign, and revoke attributes. Users can sign other user's attributes, gradually building a web-of-trust where users vouch for each others' identity attributes. As with uPort, our approach differs in that it provides a managed model that can provides additional assurances on claims.

Ethereum Improvement Proposal 725: Ethereum Improvement Proposal 725 [Vog17] (EIP-725) defines a smart contracts based identity management framework where each identity account is a separate smart contract. It supports self-attested claims and third party attestation. EIP-725 is augmented by EIP-735 [Vog], which specifies standard functions for managing claims and is supported by the ERC-725 Alliance [ERC]. An online ERC-725

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

DApp demonstration is available [0RI]. Our approach has similar capabilities but does not require every user and issuer of claims to have their own smart contract; ours is also a hierarchical managed model.

Sovrin: Sovrin is 'a protocol and token for self-sovereign identity and decentralized trust' [Sov]. Its goal is to replace the need for PKIs and to create a Domain Name System (DNS) type system for looking up public keys to be used for identity management purposes through building a custom blockchain system. It is a permissioned based cryptocurrency with no consensus protocol, thus it has centralized ownership of the tokens. The managing Sovrin foundation must approve all nodes managing the blocks but is appealing for community involvement in running nodes. The token is a cryptocurrency so that value can be exchanged along with supporting identity transactions. Our approach differs in that it doesn't require its own blockchain or cryptocurrency and can be executed on top of any smart contract system.

Decentralized Identity Foundation: The Decentralized Identity Foundation (DIF)is a large partnership with the stated goal of building an open source decentralized identity ecosystem [Fou18]. The primary focus is on high level framework, organizational issues, and standards. DIF plans to develop a broad, standards based ecosystem that supports a range of different implementations.

Other Related Work: There are many other FIM related blockchain projects that cannot be referenced here due to space limitations. For the majority of them, the design details are unavailable or are in constant flux due to the nascent nature of this market.

10 Conclusions

We have demonstrated that it is possible to design a FIM system that enables direct user to RP authentication and attribute transfer without the involvement of a third party. We implemented this using a smart contract and identified the advantages of taking such an approach. We note that user to RP interactions do not require transactions with the contract, making them fast, free, and private.

Our approach provides strong user self-sovereignty so that only the user can view and share their attribute data. However it is a managed system, intentionally not fully self-sovereignty as with the cited related work to prevent users from unilaterally changing their own identity and to provide greater validation of attribute providers. Our limits on self-sovereignty also enable the IDMS to provide authoritative and consistent data about users and participating organizations. Our approach is thus suitable for a large organization to provide identity management services to its constituents (e.g., a government). Once established by a large entity, other organizations may leverage the IDMS to provide attributes to their users and gain the ability to identify and authorize users (but only with user permission). If the owner of the contract opens up the system to many account managers and attribute managers, this will create a powerful identity management ecosystem (as opposed to being a service only for a particular purpose).

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

Bibliography

[0RI] 0RIGIN. *0RIGIN Protocol Demo*. URL: https://demo.originprotocol.com (visited on 02/05/2019).

[AlB17] Mustafa Al-Bassam. "SCPKI: a smart contract-based PKI and identity system". In: *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts*. ACM. 2017, pp. 35–40.

[ERC] ERC-725 Alliance. *ERC-725 Ethereum Identity Standard*. ERC-725 Alliance. URL: http://erc725alliance.org (visited on 02/05/2019).

[Eth] Ethereum Foundation. *Ethereum*. URL: https://www.ethereum.org/ (visited on 02/05/2019).

[Fou18] Decentralized Identity Foundation. *DIF*. 2018. URL: http://identity.foundation (visited on 04/20/2018).

[Gar95] Simson Garfinkel. *PGP: pretty good privacy*. Ö'Reilly Media, Inc.", 1995.

[Lun+17] Christian Lundkvist et al. "Uport: A platform for self-sovereign identity". 2017. URL: https://whitepaper.%20uport.%20me/uPort%5C_%20whitepaper%5C_DRAFT20170221.pdf.

[Sov] Sovrin Foundation. *Sovrin: A Protocol and Token for Self-Sovereign Identity and Decentralized Trust*. URL: https://sovrin.org/wp-content/uploads/2018/03/Sovrin-Protocol-and-Token-White-Paper.pdf (visited on 02/05/2019).

[TA18] David Temoshok and Christine Abruzzi. *NISTIR 8149 Developing Trust Frameworks to Support Identity Federations*. Tech. rep. National Institue of Standards and Technology, 2018. DOI: 10.6028/NIST.IR.8149.

[uPo18] uPort. *uPort Developers*. uPort. 2018. URL: http://developer.uport.me/overview.html (visited on 04/20/2018).

[Vog] Fabian Vogelsteller. *ERC: Claim Holder # 735*. GitHub. URL: https://github.com/ethereum/EIPs/issues/735 (visited on 02/05/2019).

[Vog17] Fabian Vogelsteller. *EIP 725: Proxy Identity*. Ethereum Foundation. Oct. 2, 2017. URL: https://eips.ethereum.org/EIPS/eip-725 (visited on 02/06/2019).

[Yag+18] Dylan Yaga et al. *NISTIR 8202 Blockchain Technology Overview*. Tech. rep. National Institue of Standards and Technology, 2018. URL: https://csrc.nist.gov/publications/detail/nistir/8202/draft.

Mell, Peter; Dray Jr., James; Shook, James.
"Smart Contract Federated Identity Management without Third Party Authentication Services."
Paper presented at Open Identity Summit 2019, Garmisch-Partenkirchen, Germany. March 28, 2019 - March 29, 2019.

PUBLIC SAFETY COMMUNICATION USER NEEDS: VOICES OF FIRST RESPONDERS

Shaneé Dawkins, Kristen Greene, Michelle Steves, Mary Theofanos, Yee-Yin Choong, Susanne Furman
National Institute of Standards of Technology, 100 Bureau Drive, Gaithersburg, MD 20899

Sandra Spickard Prettyman
Culture Catalyst, LLC

The public safety community is transitioning from land mobile radios to a communications technology ecosystem including a variety of broadband data sharing platforms. Successful deployment and adoption of new communications technology relies on efficient and effective user interfaces based on understanding first responder needs, requirements, and contexts of use; human factors research is needed to examine these factors. As such, this paper presents initial qualitative research results via semi-structured interviews with 133 first responders across the U.S. While there are similarities across disciplines, results show there is no easy "one size fits all" communications technology solution. To facilitate trust in new communications technology, solutions must be dependable, easy to use for first responders, and meet their communication needs through the application of user-centered design principles. During this shift in public safety communications technology, the time is now to leverage existing human factors expertise to influence emerging technology for public safety.

INTRODUCTION

The public safety community performs the vital mission of protecting lives and property – from day-to-day operations to out-of-the-ordinary situations. Yet, the public safety community faces significant communications challenges including interoperability and network capacity, coverage, and service. To help address such challenges, a Nationwide Public Safety Broadband Network (NPSBN) is in development for public safety to take advantage of new technological innovations and enhance their communications and information sharing. The NPSBN will enable law enforcement officers, firefighters and emergency medical services providers to send data, images, video, and location information in real-time. These new capabilities should help first responders perform their life-saving mission more safely, efficiently, and effectively.

Traditionally in the public safety communications domain, systems have been tested and measured according to network factors such as capacity, coverage, service, and other public safety-grade features. As technologies mature, it is critical that all system factors be considered, including human factors. Careful consideration of user needs will enable significant improvements in overall public safety mission delivery, much more so than technology advancements alone will achieve.

As with many technological paradigm shifts, new opportunities also present new research and development (R&D) challenges. To facilitate this new R&D, the Public Safety Communications Research (PSCR) program was established at the National Institute of Standards and Technology (NIST). PSCR has recognized the need for usability research and enhanced user interfaces as one of several major priority research areas necessary to support new public safety communications technology. PSCR believes that, in order for first responders to execute operations successfully, technology must support their ability to efficiently and effectively complete their tasks without interference – success requires a "sound understanding of the user, their requirements, and the inherent features that make a system usable" (PSCR, 2018).

The recognition of the role that user interfaces play and the call for enhanced interfaces by PSCR presents a unique opportunity for human factors researchers and practitioners. The time is now to leverage this opportunity and contribute human factors expertise to influence new products and services for the public safety communications domain.

There have been notable pockets of human factors research in the public safety space—much of it by HFES researchers (e.g., Timmons & Hutchins, 2006; Lai, Entin, Dierks, Raemer, & Simon, 2004). However, public safety has not typically benefited from the same widespread human factors attention that other domains such as the military and aviation have received. With the upcoming technological shift and potential funding in the public safety communications space, the human factors community is well-poised to offer significant contributions and lessons learned from other relevant fields.

Unfortunately, research and development communities often propose communication and technology solutions for first responders without having a full understanding related to the characteristics of their work and the problems they face. However, understanding their user needs is a crucial first step towards successful technology design, deployment and adoption. There are roughly 4 million public safety workers in the U.S., composed of firefighters (FF), emergency medical services (EMS), law enforcement (LE), and 911 center communications (COMMS) personnel. Although it is a significant undertaking, this project set out to understand the wide range of first responders, their tasks, and contexts of first responder work. The research questions guiding this effort were: How do public safety personnel describe the context of their work, including roles and responsibilities as well as process and flow? How do public safety personnel describe

Dawkins, Shanee; Greene, Kristen; Steves, Michelle; Theofanos, Mary; Choong, Yee-Yin; Furman, Susanne; Prettyman, Sandra.
"Public Safety Communication User Needs: Voices of First Responders."
Paper presented at 2018 Human Factors and Ergonomics Society (HFES) International Annual Meeting, Philadelphia, PA, United States.
October 1, 2018 - October 5, 2018.

their communication and technology needs related to work? What do public safety personnel believe is working or not working in their current operational environment?

A goal of this research effort was to engage directly with first responders to understand their current user experience and what they need in order to communicate efficiently and effectively. Engaging with first responders captures their voices so that they become audible to a broader community.

METHOD

There are two phases in this project's data collection; they are distinguished by the types of data collected. The first phase, the qualitative component, focused on interviews with 133 first responders across the U.S. The second phase will utilize a quantitative survey instrument to collect data to confirm and expand on the needs and problems related to communication and technology identified in the qualitative data. The qualitative and quantitative phases complement each other, providing a holistic view of first responders, their work, beliefs, and needs related to communications technology. The initial results for the FF, EMS, and LE disciplines of the qualitative component are described in this paper. (Note that few COMMS are included here, as they were part of a separate data collection effort.)

Qualitative research is iterative in nature and focuses on the importance of participants' perspectives throughout the research process. Our research process consistently returned to our research questions to inform elements of the process: instrument development, data collection, and data analysis. Further, data collection and initial data analysis were conducted in tandem and occurred iteratively.

Instrument Development

The research team developed a semi-structured interview protocol. Pilot interviews were conducted with several first responders in each discipline to determine face and construct validity as well as assess language appropriateness for the user population. The pilot interviews demonstrated that a generalized instrument worked well across all disciplines. The final protocol included questions on work-related tasks, relationships, and communication and technology tools, and a short demographic form. The demographic questions focused on gender, age, years of service, and participants' ease and comfort with technology.

Sampling Strategy and Participants

Sampling strategy. There is a wide range of different types of first responders with different roles and responsibilities, as well as different communications and technology needs. The initial sampling strategy focused on FF, EMS, and LE in urban, suburban, and rural locations. In addition to discipline and location representation, the sampling strategy addressed first responders with various levels of experience as a responder, from rookie to senior, as well as responders from multiple jurisdictional agency levels, and a mix of station types, both volunteer, career, and combined.

Participants. 133 participants were interviewed in 105 interview sessions. Table 1 shows the distribution of participants by discipline and location type.

Table 1: Participant distribution by discipline and location

	FF	EMS	LE	COMMS	PS*	*Total*
Urban	35	11	26	1		*73*
Suburban	25	6	12	1	2	*46*
Rural	5	3	5	1		*14*
Total	*65*	*20*	*43*	*3*	*2*	*133*

*PS are cross-trained in FF, EMS, and LE.

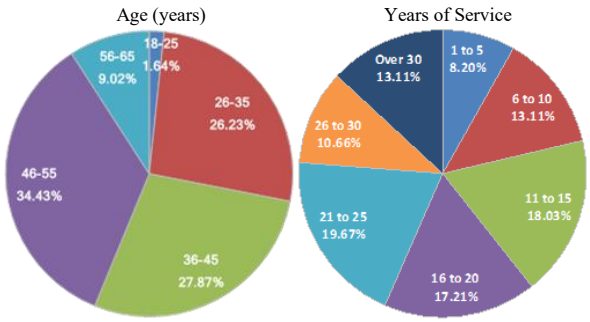


Figure 1. Participant Demographics

Figure 1 shows basic demographic characteristics. For these participants, their years of service ranged from less than a year in the field to 40 years of service (mean=18.89, *SD*=9.38). With respect to gender, there were very few women in our interview sample, just 9.84 %; however, this is representative of the first responder community in general. Women make up approximately 13 % of LE (U.S. Department of Justice, 2013) and less than 5 % of FF (National Fire Protection Association, 2018). Note that the demographics in Figure 1 are from *n*=122 first responders, as some participants did not complete all questions on the demographic questionnaire.

Data Collection

Most interviews took place in the workplace (a police, fire, or EMS station) typically in either a group gathering area, a private office, or conference room. Some interview sessions had more than one participant. Each participant was provided with a copy of the study information sheet and verbally given a summary of its content. Participants were asked for permission to audio record the session.

The data consist of interview transcripts of the recordings, demographics, field notes, and analytic memos. All interview recordings (a total of 5 627 minutes) were transcribed by an external transcription service and the transcripts form the major dataset for analysis (1 807 pages). In addition, research team members wrote field notes related to interviews they conducted, that served as additional data for analysis.

Dawkins, Shanee; Greene, Kristen; Steves, Michelle; Theofanos, Mary; Choong, Yee-Yin; Furman, Susanne; Prettyman, Sandra.
"Public Safety Communication User Needs: Voices of First Responders."
Paper presented at 2018 Human Factors and Ergonomics Society (HFES) International Annual Meeting, Philadelphia, PA, United States.
October 1, 2018 - October 5, 2018.

Data Analysis

Data analysis involved both individual and research team coding and analysis sessions. Qualitative data analysis included coding, data extraction, and analytic memoing. Coding is a process of tagging data that allows for data reduction at the start of the analysis process. In the coding sessions, an initial code list was constructed and revised as the data were explored more fully. In these sessions findings were also discussed, ideas and concepts in and about the data were explored, and concepts and variables were ultimately identified to address as part of our analysis. Finally, data extraction was performed, which is the process of pulling (extracting) all data associated with a particular code from the source data to help identify relationships and themes that might exist across codes during analysis.

In the analysis sessions, the research team explored initial ideas about the data and the codes, and began to identify any relationships and themes. Analysis included thematic, negative case, values, and descriptive exploration of the data and the codes (Saldaña, 2013). Team members used analytic memoing to document the relationships of the data and the codes. The iterative process of reviewing the data and codes, the full data set and extracted files, facilitated the identification of themes, trends, outliers, and an overall impression and understanding of the data.

RESULTS

This section presents the results of the initial analysis of the qualitative interview data, including how first responders' work influences their communication and technology needs, the challenges they face using their current technology, and their vital need for usable solutions. Participant quotes are presented to serve as exemplars of key concepts, ideas, and themes identified in the analysis rather than as just singular examples of data. All participant responses in blue text are verbatim and come directly from participant transcripts. The participant responses are followed by a notation that is comprised of three parts: discipline (FF, EMS, LE, COMMS, PS); city type (Urban=U, Suburban=S, Rural=R); and interview number. Thus (FF-R-009) refers to a FF interview, from a rural location, who was fire interview number 009.

Public Safety Communication and Technology Needs

First responders in fire, law enforcement, and EMS require unique skillsets to respond to incidents in their respective disciplines. They are responsible for handling the most extreme incidents, utilizing the highest levels of specialized training, as well as the more routine day-to-day tasks. In their work environments, first responders are expected to know it all:

> Pretty much, we're going to go help anybody that needs help, whether it's medical, fire, anything like that. If they need something, just call the fire department, we'll come and help them… So if it was really coming down to, "What do we do?" just be anything for the citizens we work for, pretty much. (FF-R-009)

> We're multi-faceted. We're teachers, doctors, nurses, medics, moms, dads, coaches, counselors is a big one, mental health specialists, which is what we get a lot of in [city redacted]. We're jack of all trades. We do everything. Report takers, problem solvers, crime fighters. I mean, we do everything in [city redacted]. (LE-U-013)

When responding to incidents, first responders have to expect the unexpected. While the extreme range of incident types in their work is overwhelmingly consistent across all three disciplines, first responders' communication and information needs and practices during incident response have very distinct differences. These differences heavily depend on their discipline, their position, and especially their role in the chain of command.

The differences among and within the various disciplines of FF, LE, and EMS imply that there is no easy "one size fits all" communications technology and data solution. Participant responses clearly show that not all first responders need access to all types of communication tools, nor to the same communication tools—everyone does not need everything. However, there are some important similarities across disciplines; for instance, FF, EMS, and LE all need to know the location and nature of incidents, and traffic patterns while en route to a location. Despite these similarities across disciplines, it is critical that technology developers and data providers know that even within a single discipline, communications technology and information needs differ based on both individual first responder roles as well as the scale and nature of the incident to which they are responding.

For FF, incident commanders need a much more holistic view of the incident in order to monitor and direct all teams involved, whereas the firefighters under their command are often completing very specific tasks and communicating only with their immediate crew. Likewise, information and communication needs for a single-family home structure fire differ from a high-rise fire or a large-scale hazmat incident. For EMS, information and communication needs of an EMS squad supervisor responsible for directing multiple crews are different from an individual paramedic and his/her partner. Coordinating and providing patient care for a mass casualty incident (MCI) is different than dealing with a single cardiac arrest patient. In LE, information and communication needs for a single patrol officer during a simple stop for a traffic violation are very different than those of an incident commander in charge of coordinating police response to an ongoing active shooter event. Shorter duration incidents require different information and communication needs than more extended incidents, such as active shooters, public protests and sporting events. Across these disciplines, the challenges in using communications technology are heightened due to first responders' unique environments, tasks, and needs.

User-Centered Design of Public Safety Communications Technology

As the human factors community is well aware, the design of new technology should focus on the user rather than be designed in a vacuum, absent of user needs. To have a positive impact on the work of first responders, meaningful improvements to their current technology and research and developments of new technology must be designed with and for them.

> The idea of [a button used only for] emergency alerting on radios is absolute crap. It's a theory and a concept that was created in an air-conditioned room on a whiteboard, but when you're scared to death, you're going to do what you do 99.9 % of the time… hit the side button [used for normal radio transmissions]. (FF-R-008)

First responders spoke of what is working or not working in their current operational environments. Universally, participants all wanted better, faster, and cheaper technology. They also emphasized the needs to improve their current technology, reduce unintended consequences, lower product/service costs, and make technology easier to use. The user-identified needs and requirements provide a blueprint for the public safety communications R&D community to develop solutions for solving the "right problems." From the first responders' interviews, six user-centered governing principles emerged that should guide all development of public safety technology.

1) Improve current technology. Designers should make what first responders currently have better, more affordable, and more reliable. For example, better radios – coverage, durability, clarity; better microphones and cords. It is not necessarily new technology that first responders want, but the improvement of current technology that they believe is most important.

> let's slow our horses a little bit, and let's back up and… Instead of introducing all this extra new stuff let's, one, make sure what we have actually works better. And then, two, let's not rely on it so much. (FF-U-042)

2) Reduce unintended consequences. Develop technology that does not take away first responders' attention from their primary tasks–causing distraction, loss of situational awareness, cognitive overload, and over-reliance on technology. Consider the social, political, policy, and legal implications of technology.

> With all the different [technology] functions, it makes actually seeing what's going on in the neighborhood harder. And somebody's looking at this box to tell them what's going on as opposed to actually looking at the surroundings and figuring out what's going on. (LE-U-024)

3) Recognize 'one size does not fit all'. While standardization is critical for consistency, compatibility and quality, technology development must accommodate a variety of different public safety needs–across disciplines, personnel, departments, districts, contexts of use–all requiring adaptability and configurability.

> So that's the challenge [for developers]. Whatever you come out with, it's not going to be one size fits all. (EMS-U-001)

4) Minimize 'technology for technology's sake'. Develop technology with and for first responders driven by their needs, requirements, and contexts of use.

> I can't make any of my employees do anything. Okay. They're here 24 hours a day. I've done their job. It's not easy. If you throw all kinds of harder stuff to make their job harder on top of it, it's not going to work. I mean, I can put all of the sanctions and rules and everything I want on it, but I have to motivate people to want to use this technology and show them the advantage of using it… We can post statistics that show us what we're really doing, how it's useful. But if it's not [useful] to them, what's in it for them? (EMS-R-008)

5) Lower product/service costs. Develop technology at price points that departments can afford to purchase and maintain to reduce monetary barriers.

> [Technology] has to be affordable, and that's the challenge. Of course, they're loosely related. I mean, there are companies out there that sell all this stuff, but it's never achievable for us. We'll never be able to spend $10 000 on a radio. We have a hard enough time spending-- right now, I mean, our radios are costing almost 4 grand for radio. And that's why we have older radios because we can't afford the new stuff. (FF-R-019)

6) Require usable technology. Develop 'Fisher-Price' solutions – simple, easy to use, light, fast, and not disruptive. Technology should make it easy to do the right thing, hard to do the wrong thing, and easy to recover when the wrong thing happens.

> But when you're in a dynamic environment, you need relatively simple what I call "Fisher-Price technology." Big shapes, big buttons, colors, things like that so that I don't have to scroll down menus and things like that… (FF-S-035)
>
> In the police world, if you want somebody to use something, it has to be simple. The more complicated it is, it's very seldom getting used. (LE-R-001)

Participants were not opposed to technology, but they want technology that makes sense to them and makes their work easier to accomplish. They don't want technology to sever and replace the human connection they see as so important. Technology must also work with first responders' other equipment and tools, and be affordable. Adhering to these guiding principles will promote first responders' trust with new technology, a requirement for successful adoption.

Trust in Public Safety Communications Technology

Early in the data analysis, the concept of trust emerged as an overarching element in public safety communications technology. The relationship of trust to many of the problems identified by the first responder participants was prevalent in the data. Although it was not specifically asked during the interviews, trust cut across the data, irrespective of discipline, geography, city type, rank, age, years of service, and other variables.

Trust is built over time and requires good experiences. Unfortunately, many participants' experiences with new technology have predisposed them not to trust technology. As a result, often participants did not see the need for new technology.

> I mean, the big thing is everything we use, I mean, we don't have time to mess with it, or tweak it, or play with it. It has to work the first time, every time, or people will just stop using it. They will just refuse to use it and go back to the old way of talking on the radio. (EMS-U-003)

Users' first impressions of a new technology heavily influences trust of that technology, adoption, and continued use. Building trust requires that the technology development community 'solve the right problems' – problems identified by first responders that impact their work.

DISCUSSION AND CONCLUSIONS

The voices of first responders should be at the forefront when considering the design, development, and adoption of public safety communications technology. According to participant responses collected in our interviews, researchers and developers should focus on technology that facilitates first responders' primary tasks and improves the user experience. During the interviews, participants noted the huge costs that occurred when technology was mandated or "pushed" upon first responders. Even if new technology is adopted at the administrative level, it will be impossible to convince end users to use it if they do not see immediate and tangible benefits for themselves. The components of usability, i.e., effectiveness, efficiency, and satisfaction, (ISO, 2010) must be fulfilled in order for successful public safety technology development and deployment. For the public safety space:

- Effectiveness – how the technology will be useful in first responders' primary task of protecting lives and property while preserving or enhancing situational awareness,
- Efficiency – how the technology will be easy to use and save first responders' time,
- Satisfaction – how the technology will promote first responders' comfort and confidence in use.

When designing for usability, it is important to consider the finding that communication and technology needs differ by first responders' roles and operating environments. This finding—that user characteristics and contexts of use influence user needs—is not unique to the public safety domain. However, it is often the case that technology designers in this space may assume that similarities among first responders outweigh their differences. Additionally, although many assume public safety is very similar to the military, first responders often have different training and experiences than do military personnel. For example, in the fire service, where 70 % of first responders are volunteer (Haynes & Stein, 2017), military-grade technology is neither designed to be cost effective, nor built to withstand the extreme environment of a fire ground (Hamins, et. al., 2015).

Differences such as those between career and volunteer in the fire service, as well as age, rank, position, or city type in the other public safety disciplines, are avenues for future exploration of the existing qualitative data results. Together with the quantitative survey results in phase two, this analysis will provide a more holistic view of challenges in the field, as well as implications for public safety technology designers and evaluators.

As human factors researchers, we need to challenge the assumption that, with new technology, public safety first responders will simply be able to 'communicate' with one another as long as their radios or devices are on the same network. There are many other non-technological factors involved: for example, differences in standard operating procedures (SOPs), communication styles, and whether first responders have trained and worked together before.

Just because devices can 'talk to each other' on a national broadband network, does not necessarily mean that first responders can do so as easily. Trust in technology is not built in a day, and it is much easier to destroy trust than it is to build it. New communications technology must be dependable, easy to use for first responders, and meet their communication needs. With the forthcoming NPSBN, the time is now for human factors professionals to partner with public safety researchers to improve public safety communications technology.

DISCLAIMER

REFERENCES

Hamins, A., Grant, C., Bryner, N. P., Jones, A. T., & Koepke, G. H. (2015). Research Roadmap for Smart Fire Fighting. NIST Special Publication 1191. DOI: http:/dx.doi.org/10.6028/NIST.SP.1191

Haynes, H. J. G., & Stein, G. P. (2017). U.S. Fire Department Profile – 2015. *National Fire Protection Association*. https://www.nfpa.org/-/media/Files/News-and-Research/Fire-statistics/Fire-service/osfdprofile.pdf (Accessed Feb, 2018).

ISO 9241-210. (2010). *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems.*

Lai, F., Entin, E., Dierks, M., Raemer, D., & Simon, R. (2004). Designing simulation-based training scenarios for emergency medical first responders. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, 15: pp. 1670-1674.

Middle Class Tax Relief and Job Creation Act of 2012, Public Law 112–96, 126 Stat. 156, February 22, 2012. http://www.gpo.gov/fdsys/pkg/PLAW-112publ96/pdf/PLAW-112publ96.pdf (Accessed Feb, 2018)

National Fire Protection Association. Firefighting occupations by women and race. http://www.nfpa.org/News-and-Research/Fire-statistics-and-reports/Fire-statistics/The-fire-service/Administration/Firefighting-occupations-by-women-and-race (Accessed Feb, 2018)

Public Safety Communications Research. (2018). https://www.nist.gov/ctl/pscr/research-portfolios/user-interfaceuser-experience (Accessed Feb, 2018)

Saldaña, J. (2013) *The Coding Manual for Qualitative Researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Timmons, R. P., & Hutchins, S. G. (2006). Radio interoperability: There is more to it than hardware. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, 4: pp. 609-613.

U.S. Department of Justice (2013) Women in Law Enforcement, in *COPS – Community Policing Dispatch*. https://cops.usdoj.gov/html/dispatch/07-2013/women_in_law_enforcement.asp (Accessed Feb, 2018)

Dawkins, Shanee; Greene, Kristen; Steves, Michelle; Theofanos, Mary; Choong, Yee-Yin; Furman, Susanne; Prettyman, Sandra.
"Public Safety Communication User Needs: Voices of First Responders."
Paper presented at 2018 Human Factors and Ergonomics Society (HFES) International Annual Meeting, Philadelphia, PA, United States.
October 1, 2018 - October 5, 2018.

2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 24-27 September 2018, Nantes, France

PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps

Nader Moayeri, Chang Li, and Lu Shi
National Institute of Standards and Technology
Gaithersburg, MD, USA
Email: {nader.moayeri, chang.li, lu.shi}@nist.gov

***Abstract*—This paper describes the structure of the PerfLoc Prize Competition organized by the US National Institute of Standards and Technology (NIST). The Competition consisted of collecting an extensive repository of smartphone data, releasing the data to researchers across the world to develop smartphone indoor localization algorithms, allowing them to evaluate the performance of their algorithms using a NIST web portal, and rigorous, live testing of the Android apps implementing the best algorithms at NIST. The paper presents detailed performance analysis of the algorithms developed by the top 10 participants as well as the Android indoor localization app that won the first prize in PerfLoc. It also provides a comparison of PerfLoc with other ongoing, annual indoor localization competitions.**

***Index Terms*—indoor localization, smartphone sensor data, smartphone apps, Android, PerfLoc**

I. Introduction

With billions in daily use around the world, the smartphone is the unrivaled king of personal mobile devices. The navigation capabilities of the smartphone enabled by the Global Positioning System (GPS) are widely used for vehicular navigation or simply looking for a place to eat while on foot in unfamiliar surroundings in a city. Even though GPS does not work indoors, many applications are envisioned for smartphone indoor localization and navigation, such as navigating to a store in a shopping mall or a work of art in a museum. The smartphone could even be used by first responders for situation-awareness and command and control purposes while responding to emergencies inside buildings.

Android phones use the Fused Location Provider (FLP) Application Programming Interface (API) [1] developed by Google for indoor/outdoor localization. iPhones use Apple's Core Location Framework [2] for indoor/outdoor localization. A comprehensive performance evaluation of Android FLP and Apple Core Location is beyond the scope of this paper, but the indoor localization accuracy they currently provide may not be adequate for some applications. Hence, it is worthwhile to explore whether more accurate smartphone indoor localization apps can be developed. With that goal in mind, the US National Institute of Standards and Technology (NIST) created and ran the PerfLoc Prize Competition [3] from March 2017 to April 2018. The preparatory steps for the Competition, however, started in August 2015. This paper describes PerfLoc from inception to conclusion.

The rest of the paper is organized as follows. Section II describes our smartphone data collection campaign. The process we used for over-the-web performance evaluation of PerfLoc "algorithms" during the Competition Testing Period that ran from March 2017 to January 2018 is described in Section III. Section IV presents a performance analysis of the top algorithms that were developed during the Competition Testing Period. Two finalist teams were invited to NIST for live tests of their "apps". The details of that evaluation and how that finalist apps performed are presented in Section V. Related work is described in Section VI. Finally, concluding remarks are provided in Section VII.

II. Smartphone Data Collection

Our data collection campaign was carried out in two phases. The original PerfLoc data was collected in early 2016. In fall 2017, i.e., about six months into the Competition Testing Period, NIST collected additional training data sets that were released to the PerfLoc user community in response to their request for such data.

A. Original Data Collection

PerfLoc was developed for the Android Operating System (OS) only, because the iPhone platform is closed and its sensor measurements and RF data are not readily accessible. NIST used four Android phones of different brands to collect data from sensors, such as accelerometer, gyroscope, magnetometer, and barometer, as well as Wi-Fi Received Signal Strength Indicator (RSSI), the strengths of signals received from cellular base stations, and GPS fixes that were occasionally available inside buildings in early 2016. The phones used for data collection were LG G4, Motorola Nexus 6, OnePlus 2, and Samsung Galaxy S6. They were strapped to the arms of the person who collected the data as shown in Figure 1. We used different smartphones to determine whether the data collected varied significantly across different brands and how those differences would affect the performance of PerfLoc indoor localization algorithms. The total space in the four buildings was more than 30,000 m^2. The buildings were a subterranean research facility with two levels below ground level, a three-story office building, a large single-story building housing a warehouse and industrial shops, and a single-story machine shop. We refer to them as Buildings 1-4, respectively, in the rest of the paper. PerfLoc competition participants

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

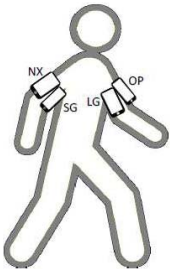


Fig. 1. Positioning of the phones on test subject's arms (LG = LG G4, NX = Motorola Nexus 6, OP = OnePlus 2, and SG = Samsung Galaxy S6)

Fig. 2. Interior of basement level in Building 1

Fig. 3. Interior of Building 4

were not told which building corresponded to which number. To the extent possible, the buildings were selected based on guidance from the international standard ISO/IEC 18305, Test and evaluation of localization and tracking systems [4], whose development NIST led during 2012-2016. Figures 2 and 3 show the interior of Buildings 1 and 4, respectively. The former shows the basement level of Building 1 that houses Heating, Ventilation, and Air Conditioning (HVAC) equipment. The building has a sub-basement level also. Building 1 did not have any Wi-Fi access points (APs) and hardly any cellular or GPS signal was ever received in that building. The relatively small number of weak Wi-Fi signals received in that building were from APs in neighboring buildings, whose locations we had not surveyed. Therefore, Building 1 was the most challenging from an RF signal availability point of view. Building 2 had a good number of Wi-Fi APs with one AP for roughly every 225 m^2 of space. The huge Building 3 had only a few Wi-Fi APs, such that in roughly half of its area no Wi-Fi signal could be received at all. Building 4 was the smallest of the four and it had a few Wi-Fi APs. One could receive Wi-Fi signals everywhere in that building, but from a couple of APs only. We installed more than 900 dots, circular floor markers of diameter 3 cm, in the four buildings and had the dot locations as well as the Wi-Fi AP locations in the buildings professionally surveyed. The dots were used as test points for the algorithms developed by PerfLoc competition participants.

We described Wi-Fi signal availability in detail, because the Wi-Fi signal is the primary information that could be used in PerfLoc to mitigate the drift inherent in localization based on an Inertial Measurement Unit (IMU). (The GPS signal was mostly unavailable in the buildings and localization based on cellular signals is notoriously inaccurate [5].) NIST did not systematically collect Wi-Fi fingerprints in the four buildings, but it provided the 3D coordinates of all Wi-Fi APs and their radio MAC addresses (BSSIDs). This is an important distinction between PerfLoc apps and Android FLP or Apple Core Location. The latter two do not have the Wi-Fi AP locations available to them, but Google and Apple collect a lot of data when a smartphone user allows them to track his/her location. Presumably, Google and Apple have large repositories of Wi-Fi data in buildings. NIST is not privy to how Google and Apple use that information.

We collected timestamped training and test data sets. The former comes with ground truth location at each timestamp so that PerfLoc competition participants could assess the accuracy of their indoor localization algorithms. We collected data over 34 test & evaluation (T&E) scenarios in the four buildings, including one training set for each building. This resulted in roughly 15.6 hours of data collected with each phone, which makes PerfLoc a very comprehensive repository of smartphone data. Each T&E scenario involves following a pre-determined course in a building using one or more mobility modes. The mobility modes used were walking normally, walking normally but pausing for 3 s at each dot visited, running, walking backwards, walking sideways (sidestepping), crawling on the floor, using an elevator instead of stairs to change floors, and placing the four phones on the bed of a cart that we pushed around in buildings. For each T&E scenario, we provided the 3D coordinates of the starting point and the initial direction of motion (E, N, W, or S).

We also provided the footprint of each building to allow PerfLoc algorithms to reject location estimates that fell outside the footprint. Specifically, we provided the coordinates of all corners of each building in counterclockwise order.

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

More details about our original data collection campaign can be found in [6], where we presented statistical analysis of the collected data such as how fast various sensors could be sampled, how periodic the sensor samples were (statistics of inter-sample times), whether different phones saw the same number of Wi-Fi APs, and the correlation of accelerometer data to the motion the phone experienced.

B. Supplemental Data Collection

We had deliberately not specified in the original PerfLoc data when a transition takes place from one mobility mode to another. The PerfLoc competition participants asked NIST to provide additional training data sets that they could use to develop models for various mobility modes as well as smartphone attitude estimation, i.e. in which 3D direction the phone is facing. NIST deployed an additional 386 dots on the floor in a building, one yard (3 floor tiles) apart from each other, and collected several training data sets using the original four phones mentioned above as well as a Google Pixel XL phone that we later used for live tests of the PerfLoc finalist apps. When collecting data with the four original phones, we attached them to the arms of the person who collected the data as shown in Figure 1 or put them on a cart. In case of the Google Pixel XL phone, the person collecting data held it in one hand in front of his/her chest, just like how most people interact with a smartphone. The cart scenario was the same as before. Just to provide some detail, in the scenario that involved crawling on the floor, the person walked for a while, then started crawling on the floor at a time that we specified in the meta data for the scenario, and finally started to walk again at a time specified in the meta data. Whether walking or crawling on the floor, we provided the timestamps for each dot on the course, ground truth location of each dot, and of course all sensor measurements made by the phone(s) on a continuous basis. In the data set for attitude estimation, we put the phone on the floor in one of six ways that we specified in our meta data and the timestamp for visiting each dot and its ground truth location. Attitude estimation plays an important role in indoor localization, because the accelerometer, gyroscope, and magnetometer in the phone measure respective sensor values with respect to the reference coordinate system of the phone, but the phone's orientation changes continuously as a result of the body motion of the person who collected the data. Just like the original PerfLoc data, the supplemental data is available on the PerfLoc website [3] along with detailed descriptions of the data.

III. Offline Evaluation of PerfLoc Algorithms

NIST created an over-the-web performance evaluation capability for PerfLoc algorithms to provide instant objective feedback to PerfLoc competition participants on how good their algorithms were and to give them the opportunity to make their algorithms better. We developed a web portal where a PerfLoc competition participant could upload the location estimates generated by his/her algorithm for the timestamps specified in each of the 30 test data sets. The web portal would then compute the spherical error 95% (SE95) [4] for each of the four buildings as well as an overall SE95 for all test data sets and report those figures instantaneously back to the PerfLoc competition participant. Competing PerfLoc algorithms were ranked and listed in a leaderboard published on the PerfLoc website in the ascending order of overall SE95. Location estimates generated by all four phones had to be uploaded to allow assessing the performance of a given algorithm on different phones, even though in reality most PerfLoc competition participants used the measurements made by "all" four phones to estimate the location at time instances of interest and then uploaded the same location estimates for all four phones! That aspect was unfortunate, because it prevented NIST from quantifying the differences between the four phones in terms of localization accuracy achieved by the same algorithm.

A key aspect of designing the performance evaluation portal was to incorporate mechanisms to protect the integrity of the method used to rank competing algorithms and to prevent any PerfLoc competition participant from gaming the system and achieving a bogus performance. Specifically, NIST had to prevent PerfLoc competition participants from algorithmically computing the locations of our dots (test points) based on the numerical feedback they were getting from the PerfLoc performance evaluation portal. We realized that the system would be vulnerable to such abuse if we reported the mean of magnitude of 3D localization error, i.e. the difference between the ground truth location of a dot and the estimate for that location provided by the PerfLoc algorithm under test. It is possible to precisely compute a dot location by three uses of the performance evaluation portal and then solving a not-so-difficult nonlinear system of equations. Consequently, we chose not to use the mean of magnitude of 3D error vector as our performance metric, even though all other indoor localization competitions use that metric. Instead, we selected SE95 as our performance metric. We realize that even SE95 is vulnerable to abuse, but the algorithm to compute dot locations based on SE95 feedback is much more complicated than the corresponding algorithm for the mean of magnitude of 3D error vector.

Other steps we took to make it harder to abuse the performance evaluation portal was to ask each team to pledge they would create only one user account. We then limited the number of times a team could use the portal to three uploads per day and a total of 150 uploads during the Competition Testing Period. In reality, we did not have an assured way of guaranteeing that a team would not create more than one user account and there was one instance of abuse of that clause in the competition rules that led to disqualification of one team late into the Competition Testing Period. Overall, the mechanisms NIST developed and implemented in the PerfLoc Prize Competition were a compromise between preventing abuse of the system and ease of joining the competition and using the system.

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

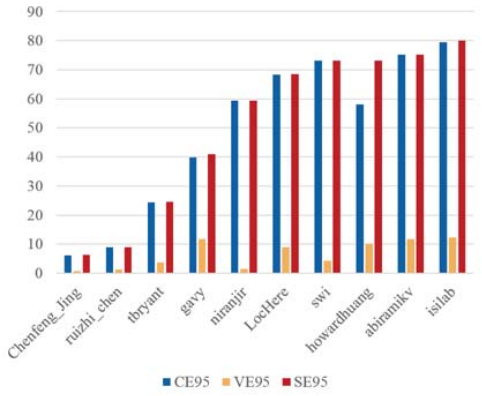


Fig. 4. CE95, VE95 and SE95 of the top 10 teams

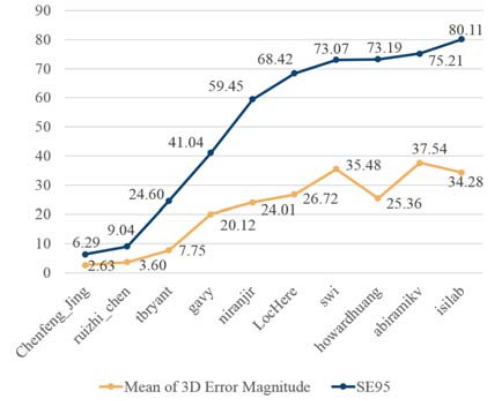


Fig. 5. Comparison of mean of 3D error magnitude and SE95 for the top 10 teams

IV. Offline Performance of PerfLoc Algorithms

A total of 152 teams registered on the PerfLoc website, but only 16 teams uploaded location estimates on the website. We attribute this to PerfLoc's steep learning curve. There were a lot of details in the PerfLoc User Guide [3] that the participants had to master before decoding the data and beginning to use it. The Competition Testing Period, during which participants were allowed to upload location estimates, lasted for almost ten months and it closed in January 2018. The overall SE95, circular error 95% (CE95), which characterizes horizontal accuracy, and vertical error 95% (VE95) of best performance achieved by the top 10 teams computed over all 30 T&E scenarios have been depicted in Figure 4. These are the 95 percentile points on the Cumulative Distribution Functions (CDFs) of the respective error magnitudes. We observe that the vertical error is much smaller than the horizontal error, and hence the 3D error is dominated by the horizontal error. The figure also shows the usernames of the top 10 teams.

Figure 5 compares the overall SE95 and mean of magnitude of 3D error achieved by various teams. It shows that the means are considerably smaller than the SE95s. The top team achieved an overall SE95 of 6.29 m and mean of magnitude of 3D error of 2.63 m.

Table I shows the SE95 and mean of magnitude of 3D error achieved by the top 10 teams in different buildings. Aside from the top two teams who have achieved roughly the same localization accuracy in all four buildings, most of the remaining teams have achieved better performance in Buildings 2 and 4 than in Buildings 1 and 3. This is explained by the fact that Buildings 2 and 4 have better Wi-Fi coverage than Buildings 1 and 3.

Table II shows the VE95 and mean of magnitude of vertical error achieved by the top 10 teams in different buildings. It shows that Buildings 3-4 results are generally better than Buildings 1-2 results. This is due to the fact that while 3-4 are single-story buildings, 1-2 have multiple floors.

V. Live Testing of PerfLoc Apps

At the conclusion of the Competition Testing Period in January 2018 and according to the published PerfLoc Competition Rules, NIST invited teams ranked 2-4 on the PerfLoc Competition Leaderboard for live testing of their apps at NIST. (The top-ranked team was not invited, because they were not eligible to receive cash prizes.) Team #4 declined the invitation. Teams #2 and #3 accepted the invitation and traveled to NIST for live testing of their apps on April 26-27, 2018. Unfortunately, Team #3 was not able to get its app to function at all and dropped out of the race. Therefore, Team #2 (with username ruizhi_chen) from Wuhan University in China, being the only team that managed to go through the suite of tests administered by NIST, was declared the winner of the PerfLoc Prize Competition.

A. Purpose and Structure of Live Tests

The purpose of the live tests were threefold. First, NIST wished to ascertain that any finalist PerfLoc algorithm could be implemented as an Android app. NIST could not determine from the offline performance results whether an algorithm would need to be run on a super computer and/or would take days to generate location estimates. Second, it was important to find out how a finalist app would fare in a blind, live test and how that would compare to the offline performance of the same app/algorithm. By blind test we mean testing an app in a building that the finalist knew nothing about until the test days. In other words, the finalists did not know how large the building was, how many Wi-Fi APs it had, and they did not have any training data for the building. NIST assumed that PerfLoc finalists had perfected their algorithms/apps during the Competition Testing Period and they were supposed to be ready for testing in "any" building by the live test days. This was indeed a tall order. Third, NIST wished to measure the latency of each finalist app, i.e. the time it takes for an app to provide a location estimate. Note that there is a tradeoff between latency and localization accuracy. If an algorithm uses a bit of lookahead, i.e. some future sensor and RF measurements, before generating a location estimate for the present time, its location estimates would be more accurate

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

TABLE I
SE95 AND MEAN OF 3D ERROR MAGNITUDE IN DIFFERENT BUILDINGS FOR THE TOP 10 TEAMS

Rank	Participant	SE95 Performance (m)				Mean of 3D Error Magnitude (m)			
		Bldg. 1	Bldg. 2	Bldg. 3	Bldg. 4	Bldg. 1	Bldg. 2	Bldg. 3	Bldg. 4
1	Chenfeng_Jing	5.44	7.04	6.26	5.38	2.14	3.19	2.89	2.23
2	ruizhi_chen	12.74	8.60	7.34	8.99	3.84	3.61	3.27	3.48
3	tbryant	31.92	17.78	20.93	17.91	8.44	8.17	6.83	6.30
4	gavy	42.71	24.56	37.28	26.33	27.34	14.23	17.97	13.90
5	niranjir	41.88	24.81	71.25	52.27	24.91	10.45	37.62	20.63
6	LocHere	73.12	21.83	74.55	40.29	36.88	10.42	32.56	15.74
7	swi	42.79	57.68	82.99	49.62	25.09	33.17	55.96	26.07
8	howardhuang	126.87	20.20	72.77	39.83	34.71	9.38	29.92	20.38
9	abiramikv	42.71	60.14	84.95	65.39	27.34	33.57	56.30	36.95
10	isilab	73.59	17.98	339.29	25.03	42.96	10.07	54.07	12.44

TABLE II
VE95 AND MEAN OF VERTICAL ERROR MAGNITUDE IN DIFFERENT BUILDINGS FOR THE TOP 10 TEAMS

Rank	Participant	VE95 Performance (m)				Mean of Vertical Error Magnitude (m)			
		Bldg. 1	Bldg. 2	Bldg. 3	Bldg. 4	Bldg. 1	Bldg. 2	Bldg. 3	Bldg. 4
1	Chenfeng_Jing	0.04	3.25	0.05	0.02	0.07	0.58	0.03	0.01
2	ruizhi_chen	0.58	2.81	0.03	0.18	0.09	0.72	0.01	0.17
3	tbryant	3.78	4.85	3.40	3.40	1.68	1.70	1.69	1.31
4	gavy	11.84	5.97	1.22	6.48	9.56	3.22	0.39	1.54
5	niranjir	0.05	2.29	0.03	0.02	0.10	0.98	0.01	0.01
6	LocHere	9.00	3.55	0.03	0.02	3.67	0.44	0.01	0.01
7	swi	2.39	4.30	0.02	0.01	2.41	2.66	0.01	0.01
8	howardhuang	122.45	5.98	0.61	1.04	15.64	3.73	0.51	1.02
9	abiramikv	11.84	7.30	0.02	0.24	9.56	4.47	0.01	0.23
10	isilab	12.40	3.41	9.13	0.02	6.26	0.72	0.69	0.01

Fig. 6. The building used for the live tests

compared to not using any lookahead. Perhaps a lookahead of ~2 s would be acceptable.

The finalist app was tested in a very large, tall building with about 30,000 m^2 of space and 131 Wi-Fi APs. A picture of this building is shown in Figure 6. Note that the first floor and the basement of this building are much larger than the tower part. This building by itself is as large as the four buildings used during the offline performance evaluation phase put together. The 3D coordinates and BSSIDs of the Wi-Fi APs, the building footprint, and the 3D coordinates and initial direction of motion for each of the 8 T&E scenarios used were provided to the finalist app. Unlike the offline phase, where NIST had provided smartphone sensor and RF data at sampling rates of NIST's choosing, the decision of how fast to sample various data was left to the developers of the finalist app. The T&E scenarios were (i) normal non-stop walking, (ii) normal walking with 3 s stops at each test point visited, (iii) transporting an asset on a push cart, (iv) normal walking and use of elevators, (v) normal walking with instances of leaving and reentering the building, (vi) sidestepping, (vii) walking backwards, and (viii) crawling on the floor.

B. Live Test Performance Results

The latency of the Wuhan University Team app turned out to be ~5 milliseconds.

Table III shows the performance of the app. The first observation we make is that the live test results are not nearly as good as the offline performance results. This discrepancy is due to (i) unavailability of training data sets in the live tests, (ii) real-time operation requirement of the live tests that prevented the app from post-processing and revising/improving location estimates for past test points, (iii) the building used in the live tests being much larger than the four used during offline performance evaluation, and (iv) the app being forced to use a single algorithm during the live tests as opposed to possibly using building-specific algorithms during the offline phase.

A few other observations can be made. The first two rows of Table III show that the horizontal error is much larger than the vertical error. The same observation can be made by comparing CE95 and VE95 figures. Normal walking's performance is considerably better than the overall (over all mobility

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

TABLE III
LIVE TESTING PERFORMANCE OF WUHAN UNIVERSITY TEAM APP

	Overall	Normal Walking	Cart	Sidestepping	Walking Backwards	Crawling
Mean of Magnitude of Horizontal Error	17.66	10.76	49.67	27.79	38.07	10.31
Mean of Magnitude of Vertical Error	1.71	2.09	0.72	0.57	0.20	0.34
Mean of Magnitude of 3D Error	18.16	11.43	49.68	27.81	38.07	10.32
CE95	72.84	24.86	78.92	79.17	83.10	18.56
VE95	5.20	5.22	1.16	0.95	0.47	0.57
SE95	72.84	24.86	78.92	79.17	83.10	18.56
Floor Detection Probability	70.66%	59.35%	100%	100%	100%	100%

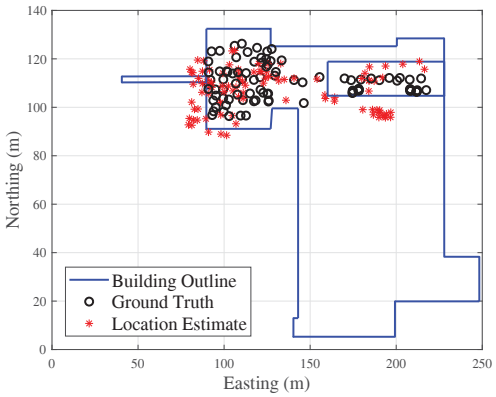


Fig. 7. Horizontal error of walking scenario with 3 s stops

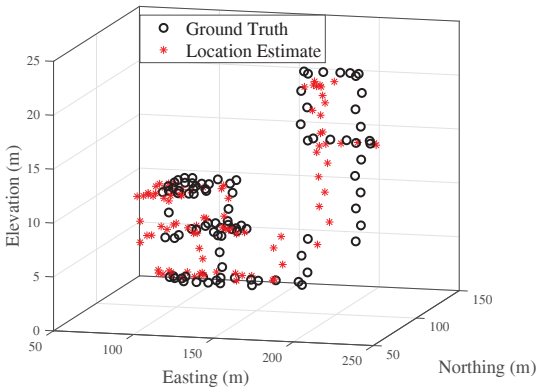


Fig. 8. Vertical error of walking scenario with 3 s stops

modes) performance. Specifically, the cart, sidestepping, and walking backwards scenarios have much worse performance than normal walking.

Figures 7 and 8, respectively, show the horizontal and vertical performance of the app in the normal walking scenario with 3 s stops at test points visited. It is hard to tell that Figure 7 represents good performance. However, at least it can be seen that the black circles and the red stars do not look like two sets separated spatially as in some figures to be introduced shortly. Figure 8 does show that the app is doing a good job of tracking the test subject's elevation.

Figures 9, 10, and 11, respectively, show the horizontal performance of the app in the cart, sidestepping, and walking backwards scenarios. It is clear that the app is not doing a good job of tracking the test subject's location. Sometimes it overestimates how far the test subject has moved and sometimes it underestimates. In all three cases, one can see a good separation of the black circles and the red stars. As shown in Table III, the vertical error achieved in these scenarios, which were carried out on the same floor of the building, is small. Therefore, no figures on vertical performance are provided for these scenarios.

Figure 12 is a plot of the magnitude of 3D error vs. time in the scenario that involved normal walking and four uses of elevators. In two cases of using the elevator marked in the figure, where we went up or down by several floors, a large increase in the magnitude of error is observed. After each such

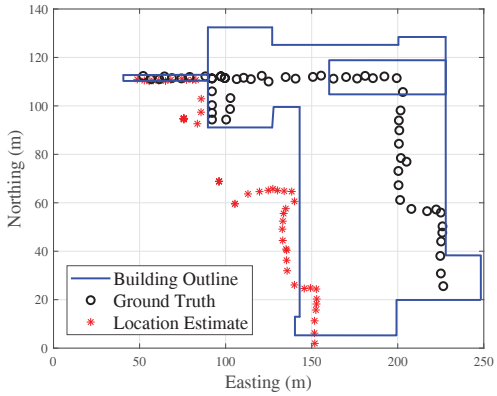


Fig. 9. Horizontal error of the cart scenario

increase, the error drops after some time, perhaps as a result of getting good location fixes from the Wi-Fi signals. There are two other cases in this scenario, where we used the elevator to go up or down by just one floor at ~330 s and ~565 s. In these cases, there is no significant change in the magnitude of error.

Figure 13 is a plot of the magnitude of 3D error vs. time in the normal walking scenario with two instances of leaving and reentering the building. We conclude that getting GPS fixes by leaving the building does not help to mitigate the

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

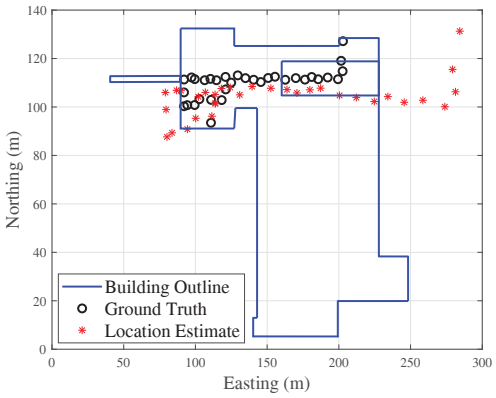


Fig. 10. Horizontal error of the sidestepping scenario

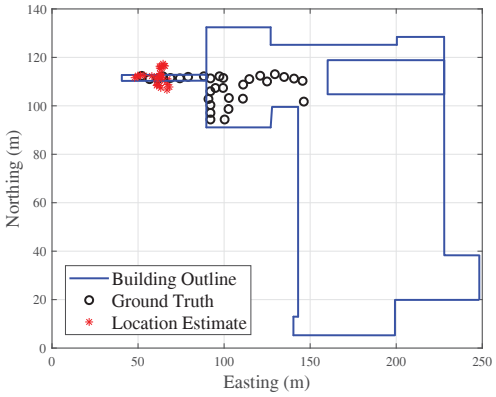


Fig. 11. Horizontal error of the walking backwards scenario

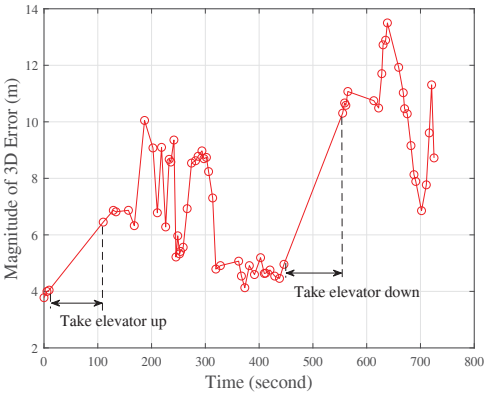


Fig. 12. Mean of 3D error magnitude vs. time for the scenario involving walking and use of elevators

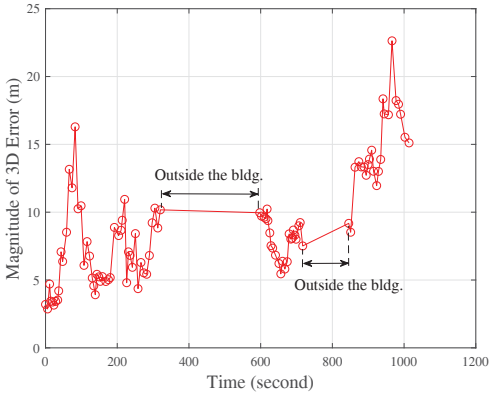


Fig. 13. Mean of 3D error magnitude vs. time for the scenario involving walking in/out of a building

drift of inertial sensors any more so than location fixes from Wi-Fi signals. Note that the error is in the range 7.5-10 m after walking outdoors and getting GPS fixes for about 4.5 minutes in the first instance of leaving the building and about 2.5 minutes in the second instance.

VI. Related Work

There are two indoor localization competitions that have been held on an annual basis for the past several years. The first one is the competition held in conjunction with the IEEE Indoor Positioning and Indoor Navigation (IPIN) Conference since 2011. These competitions are typically organized by the EvAAL (Evaluating AAL Systems through Competitive Benchmarking) Project [7], where AAL stands for Ambient Assisted Living. In addition to indoor localization, EvAAL is interested in indoor activity recognition. The second one is the Microsoft Indoor Localization Competition [8] that has been held in conjunction with the IEEE Information Processing in Sensor Networks (IPSN) Conference since 2014. The structure of these competitions may change somewhat from one year to the next. We describe the structure for the latest edition of each competition prior to the writing of this paper.

The 2018 Microsoft Indoor Localization Competition was held in two tracks, (i) 2D Track and (ii) 3D Track. The systems competing in the 2D Track used technologies such as Pedestrian Dead Reckoning (PDR), camera, and Wi-Fi fingerprinting. One system used Wi-Fi Time of Flight (ToF). The competing systems were evaluated based on mean horizontal localization error and the best system achieved 2.3 m mean error. In the 3D Track, the contestants were allowed to install up to 10 anchor nodes of their choice in the evaluation area, which was about 600 m^2. Different systems were compared based on the mean 3D error performance metric. In addition to the technologies used in the 2D Track, the systems competing in this track also used ultra wideband (UWB) ranging, sound, and ARKit [9]. One system used a phase-based localization technique. The best system in the 3D Track integrated UWB and ARKit to achieve a mean 3D error of 27 cm. Of course,

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

this is far better than the 18.16 m mean 3D localization error the winning PerfLoc app achieved in the live tests described in Section V, but such a comparison would not be fair for many reasons. First, the PerfLoc app did not have the benefit of using UWB or ARKit, granted that the use of the latter would be reasonable in a smartphone app. (In case of Android phones, one would use ARCore [10].) Second, the PerfLoc app was tested over scenarios that lasted as long as 20 minutes in a huge, tall building. Had it been tested over a 2-3 minute scenario in a small area, it would have achieved better results. Third, the PerfLoc app was tested over many different modes of mobility. Its performance for normal walking is 11.43 m, as shown in Table III. Even a comparison of PerfLoc with the 2D results of the Microsoft Indoor Localization Competition would not be fair due to the differences in the evaluation area size. Wi-Fi fingerprinting was allowed in the Microsoft Indoor Localization Competition, but ironically PDR alone was better than PDR plus fingerprinting due to the use of transmit power control by Wi-Fi APs. It is fair to say that PerfLoc played a different role than the Microsoft Indoor Localization Competition does. The latter generates a lot of interest every year as a forum for emerging techniques and solutions. PerfLoc was restricted to the Android smartphones and it used more rigorous testing.

It is difficult to compare performance results from various competitions. Even if one compares similar systems, e.g. smartphone apps that use Wi-Fi fingerprinting, one still has to take into account the differences in the evaluation areas. These issues have been addressed in detail in the international standard ISO/IEC 18305 [4]. These evaluations are site-dependent. The best that can be done is to compare several systems evaluated in the same set of buildings roughly at the same time.

The 2017 IPIN Competition was held in four tracks: (i) Smartphone-Based, (ii) PDR Positioning, (iii) Smartphone-Based (Off-Site), and (iv) PDR for Warehouse Picking (Off-Site). Track 1 was similar to the live tests of the winning PerfLoc app, but it allowed Wi-Fi fingerprinting prior to the tests (as opposed to providing Wi-Fi AP locations only in PerfLoc) and it made the detailed map of the evaluation area available to the contestants (as opposed to providing the building footprint only in PerfLoc). In Track 3, training data sets with ground truth location data were made available to the contestants to develop and fine-tune their algorithms, which were subsequently tested by the competition organizers using a data set the contestants did not have access to. Unlike PerfLoc, there were no opportunity to get any feedback on the performance of an algorithms during its development phase. Wi-Fi fingerprints were also made available to the contestants in the IPIN Track 3 Competition. The evaluation area in the 2017 IPIN Competition was about 1,500 m^2 over two floors. The best CE75 (as opposed to CE95 used in PerfLoc) in Tracks 1 and 3 were 8.8 m and 3.48 m, respectively. These performance results look better than that of the winning PerfLoc app, but one has to keep in mind that IPIN apps were tested in a smaller area, had detailed maps available to them, and could use Wi-Fi fingerprinting. In addition, CE75 is a less stringent performance metric than CE95. The evaluation area used in the 2016 IPIN Competition was even larger and covered 3 floors of a building.

VII. Conclusions

PerfLoc was a prize competition for developing smartphone indoor localization apps with the minimal requirement of having access to the locations and BSSIDs of Wi-Fi APs, if the building has Wi-Fi. This is less onerous than having to make Wi-Fi fingerprints available to apps. The Android apps developed during this competition were evaluated during an offline phase and through comprehensive live tests at NIST. The winning app achieved a mean 3D error of about 10 m when the test subject walked around the building with the smartphone in his/her hand and used stairs or elevators to change floors. The winning app did a great job of estimating on which floor of the building the person carrying the smartphone was by achieving a mean vertical error of 1.71 m. However, the app did not perform well for other modes of mobility such as sidestepping, walking backwards, or tracking the movements of an object transported on a push cart in the building. The PerfLoc Prize Competition is now closed, but the problem of developing more effective indoor localization smartphone apps with better accuracy is still very much open. The extensive repository of PerfLoc smartphone data continues to be available to researchers for doing just that.

Disclaimer

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards nd Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

References

[1] Google, "Fused Location Provider API." [Online]. Available: https://developers.google.com/location-context/fused-location-provider/
[2] Apple, "Core Location." [Online]. Available: https://developer.apple.com/documentation/corelocation
[3] US National Institute of Standards and Technology, "PerfLoc Prize Competition." [Online]. Available: https://perfloc.nist.gov
[4] "ISO/IEC 18305:2016, Information technology – Real time locating systems, Test and evaluation of localization and tracking systems." [Online]. Available: https://www.iso.org/standard/62090.html
[5] P. A. Zandbergen, "Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning," *Trans. GIS*, vol. 13, no. s1, pp. 5–25, 2009.
[6] N. Moayeri, M. Ergin, F. Lemic, V. Handziski, and A. Wolisz, "PerfLoc (Part 1): An extensive data repository for development of smartphone indoor localization apps," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Comm. (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 1–7.
[7] "Evaluating AAL Systems through Competitive Benchmarking." [Online]. Available: http://evaal.aaloa.org/
[8] "Microsoft Indoor Localization Competition." [Online]. Available: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2018/
[9] Apple Developer, "ARKit." [Online]. Available: https://developer.apple.com/arkit/
[10] Google Developers, "ARCore." [Online]. Available: https://developers.google.com/ar/discover/

Moayeri, Nader; Li, Chang; Shi, Lu.
"PerfLoc (Part 2): Performance Evaluation of the Smartphone Indoor Localization Apps."
Paper presented at 9th International Conference on Indoor Positioning and Indoor Navigation, Nantes, France. September 24, 2018 - September 27, 2018.

"We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products

Julie M. Haney[1], Mary F. Theofanos[1], Yasemin Acar[2], Sandra Spickard Prettyman[3]

[1]National Institute of Standards and Technology
{julie.haney, mary.theofanos}@nist.gov

[2]Leibniz University Hannover
acar@sec.uni-hannover.de

[3]Culture Catalyst
sspretty50@icloud.com

ABSTRACT

Cryptography is an essential component of modern computing. Unfortunately, implementing cryptography correctly is a non-trivial undertaking. Past studies have supported this observation by revealing a multitude of errors and developer pitfalls in the cryptographic implementations of software products. However, the emphasis of these studies was on individual developers; there is an obvious gap in more thoroughly understanding cryptographic development practices of organizations. To address this gap, we conducted 21 in-depth interviews of highly experienced individuals representing organizations that include cryptography in their products. Our findings suggest a security mindset not seen in other research results, demonstrated by strong organizational security culture and the deep expertise of those performing cryptographic development. This mindset, in turn, guides the careful selection of cryptographic resources and informs formal, rigorous development and testing practices. The enhanced understanding of organizational practices encourages additional research initiatives to explore variations in those implementing cryptography, which can aid in transferring lessons learned from more security-mature organizations to the broader development community through educational opportunities, tools, and other mechanisms. The findings also support past studies that suggest that the usability of cryptographic resources may be deficient, and provide additional suggestions for making these resources more accessible and usable to developers of varying skill levels.

1. INTRODUCTION

In a dynamic, threat-laden, and interconnected digital environment, cryptography protects privacy, provides for anonymity, ensures the confidentiality and integrity of communications, and safeguards sensitive information. Given the need for cryptography, there is an abundance of cryptographic algorithm and library choices for developers wishing to integrate cryptography into their products and services. However, developers often lack the expertise to navigate these choices, resulting in the introduction of security vulnerabilities [27]. A 2016 industry survey that included over 300,000 code assessments found that 39% of those applications had cryptographic problems [72]. Implementing cryptography correctly is a non-trivial undertaking.

In 1997, security expert Bruce Schneier commented on the lack of cryptographic implementation rigor and expertise at that time, asserting, "You can't make systems secure by tacking on cryptography as an afterthought. You have to know what you are doing every step of the way, from conception to installation" [61]. Past studies have supported this observation by revealing a multitude of errors in the cryptographic implementations of software products (e.g., [17–19, 42]) and the pitfalls developers encounter when including cryptography within products (e.g,. [1, 2, 48]). This body of research suggests that developers have not progressed much in the past 20 years. However, as these studies have been largely focused on individual practices outside the professional work context or on the development of mobile apps, it is unclear if these shortcomings also apply to organizational development and testing, particularly among organizations for which security and cryptography are essential components. One exploratory survey examined high-level organizational practices in cryptographic development, but lacked rich insight into actual practices and motivators behind those [31]. Clearly, there is a gap in the literature in more thoroughly understanding organizational cryptographic development practices.

To address this gap, we performed a qualitative investigation into the processes and resources that organizations employ to ensure their cryptographic products are not fraught with errors and vulnerabilities. We define the scope of cryptographic products as those implementing cryptographic algorithms or using crypto (cryptography) to perform some function. We conducted 21 in-depth interviews involving participants representing organizations that develop either a security product that uses cryptography or a non-security product that heavily relies on cryptography. Unlike previous studies, our participants were professionals who were highly experienced in cryptographic development and testing, not computer science students or developers with little cryptographic experience.

The study aimed to answer the following research questions:

Q1 What are the cryptographic development and testing practices of organizations?

USENIX Symposium on Usable Privacy and Security (SOUPS) 2018.
August 12–14, 2018, Baltimore, MD, USA.

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

Q2 What challenges, if any, do organizations encounter while developing and testing these products?

Q3 What cryptographic resources do these organizations use, and what are their reasons for choosing these?

Our findings went deeper than uncovering practices, revealing a security mindset not noted in other research results. We discovered that some organizations believe they have achieved the expertise and rigor recommended by Schneier. Compared to developer populations studied in the past, the organizations in our investigation appear to have a stronger security culture and are more mature in their cryptography and security experience. The strong security culture we observed does not appear to be linked to company size or available resources. These security mindsets permeate the entire development process as they inform judicious selection of cryptographic resources and rigorous development practices.

Our work has several contributions. To our knowledge, this is the first in-depth study to explore cryptographic development practices and security mindsets in organizations from the viewpoint of those with extensive experience in the field. While some of the practices identified in our study may be considered known best practice within the security community, our paper is novel in that there are few research studies documenting occurrences of strong security culture and development in actual practice, and none within the cryptography context. Our study provides systematic, scientific validation to the anecdotal point often made by security experts that there is no magical, one-dimensional solution to cryptographic development. Rather, good crypto is the result of a concerted effort to build expertise and implement secure development practices. The enhanced understanding encourages additional research initiatives to explore variations in those implementing cryptography. This can aid in transferring lessons learned from more security-mature organizations to the broader development community through educational opportunities, tools, and other mechanisms. Our findings also support past studies that suggest that the usability of cryptographic resources may be deficient, and provide additional suggestions for making these resources more accessible and usable to developers of varying skill levels.

2. RELATED WORK

To provide context, this section begins with a brief overview of cryptographic standards and certifications frequently referenced in our interviews. We then underpin our assertion that cryptographic development is not a trivial undertaking by summarizing past research on crypto misuse and lack of crypto resource usability. We also present an overview of prior work on lack of security mindsets and secure development practices to serve as a contrast to the more security-conscious approaches of our study organizations.

2.1 Cryptographic Standards

Cryptographic algorithm standards are developed by consensus of community stakeholders (e.g. vendors, researchers, governments) to foster compatibility, interoperability, and minimum levels of security. These can be found in formal standards documents from organizations such as Institute of Electrical and Electronics Engineers (IEEE) [35], International Organization for Standardization (ISO) [36], and the U.S. National Institute of Standards and Technology (NIST) [52]. Likely due to the U.S. locations of most of the study organizations, the participants most often mentioned cryptographic requirements issued by NIST. As perhaps the best known government standard, the Federal Information Processing Standards Publication (FIPS) 140-2 "specifies the security requirements that will be satisfied by a cryptographic module utilized within a security system protecting sensitive but unclassified information" [49]. These requirements are mandatory for cryptographic products purchased by the U.S. Government, but also are used voluntarily outside the government. There are two certification programs associated with FIPS 140-2 [50, 51]. Under these programs, vendors may submit cryptographic algorithm and module implementations for validation testing to accredited testing laboratories.

2.2 Cryptographic Misuse

Numerous studies have highlighted the difficulty developers have in correctly implementing cryptography. In 2002, Gutmann observed that security bugs were often introduced by software developers who did not understand the implications of their choices [30]. Nguyen showed that even open-source implementations under public scrutiny have cryptographic flaws [53]. Lazar performed a systematic study of 269 cryptographic vulnerabilities in the Common Vulnerabilities and Exposures (CVE) database, noting that 17% of bugs were in cryptographic libraries and the remaining 83% were in individual applications, usually due to cryptographic library misuses [40]. Georgiev et al. discovered rampant misuse of Secure Sockets Layer (SSL) in security-critical applications due to poorly designed application programming interfaces (APIs) [25]. Fahl et al. analyzed 13,500 Android apps and found that 8% were susceptible to man-in-the-middle attacks [19]. Using static analysis, Egele et al. found similar issues, observing that 88% of over 11,000 examined Android apps contained a significant error in their use of a cryptographic API [18]. Li et al. analyzed 98 apps from the Apple App Store and found 64 (65.3%) that contained cryptographic misuse flaws [42].

2.3 Usability of Cryptographic Resources

Usability is often neglected in cryptographic resources such as standards and libraries, resulting in complex solutions that provide little assistance to developers in making secure choices [27, 48]. Several research groups attempted to remedy this by developing tools, e.g., OpenCCE [4], CryptoAssistant [23] and Crypto Misuse Analyzer [64], to guide developers in choosing and integrating appropriate cryptographic methods. Others proposed more usable cryptographic libraries. Forler et al. developed libadacrypt, a cryptographic library created to be "misuse-resistant" [20]. Bernstein et al. created the Networking and Cryptography library (NaCl) [8], a cross-platform cryptographic library designed to avoid errors found in widely used cryptographic libraries like OpenSSL [55]. Acar et al. conducted a usability study of cryptographic APIs that revealed that, in addition to usable interfaces, clear documentation with code samples and support for common cryptographic tasks were important in aiding developers [1].

2.4 Security Development and Mindset

There is much to learn from an examination of secure de-

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

velopment and testing practices since even the best implemented cryptography can be subverted by the flawed implementation of another system component. McGraw advocated for security to be integrated into all aspects of the software development lifecycle [43]. Several documents, for example the Microsoft Security Development Lifecyle [34] and the System Security Engineering Capability Maturity Model (SSE-CMM) [44], formally define secure development practices. More recently, the Open Web Application Security Project (OWASP) Secure Software Development Lifecycle project is working towards providing guidelines for web and application developers [54]. However, none of these resources specifically mentions considerations for cryptography.

Not surprisingly, the implementation of formal secure development processes is not an easy task. A 2016 Veracode survey of over 350 developers indicated that organizations are prevented from fully implementing a secure development process due to a variety of challenges, including security testing causing product timeline delays, complexity in supporting legacy security processes, security standards and policies varying across the organization, and developers not consistently following secure coding practices [73]. Kanniah and Mahrin found that a variety of organizational, technical, and human factors affected implementation of secure software development practices [38]. These factors included developer skill and expertise, communication among stakeholders, and collaboration between security experts and developers.

Failures in development and testing leading to security errors appear to reflect a deficiency in a security mindset. Schneier claimed that "Security requires a particular mindset. Security professionals – at least the good ones – see the world differently" [62]. A security mindset involves being able to think like an attacker, maintaining a commitment to secure practices, and perpetuating a strong security culture.

A need for a security mindset is revealed in several studies that explored reasons why developers make security errors. For example, Xie, Lipford, and Chu identified an absence of personal responsibility for security as well as a gap between developers' understanding of security and how to implement it [76]. Xiao et al. discovered that the failure to adopt secure development tools was heavily dependent on social environments and how tool information was communicated [75]. From a testing perspective, Potter and McGraw argued that security testing is commonly misunderstood and should be more risk-based, involving an understanding of a potential attacker's mentality [58]. Bonver and Cohen agreed, noting that security testers should work closely with architects and developers to identify potential vulnerabilities, taking into account how an attacker may exploit a system [9].

3. METHODOLOGY

Between January and June 2017, we conducted 21 interviews of individuals working in organizations that develop products that use cryptography. Following rigorous, commonly accepted qualitative research methods, we continued interviewing until we reached theoretical saturation, the point at which no new themes or ideas emerged from the data [45], exceeding the minimum of 12 interviews prescribed in qualitative research best practices [29].

Our research team was multidisciplinary and consisted of a computer scientist specializing in information security and human-computer interaction, a computer scientist specializing in usability, a mathematician with research experience in usable security and privacy, and a sociologist experienced in qualitative research. Having a diverse team may improve research quality "in terms of enabling sounder methodological design, increasing rigor, and encouraging richer conceptual analysis and interpretation" [6].

The study was approved by our institutional review board. Prior to the interviews, participants were informed of the purpose of the study and how their data would be used and protected. Interview data were collected and recorded without personal identifiers and not linked back to the participants or organizations. Interviews were assigned an identifier (e.g., C08) used for all associated data in the study.

3.1 Recruitment

To ensure that we could explore different perspectives within the cryptographic product space, our sampling frame consisted of individuals who had organizational experience designing, developing, or testing products that use cryptography or who were knowledgeable about and had played a key role in these activities (e.g., managers of teams that performed these tasks). We utilized a combination of purposeful and convenience sampling strategies, which are widely employed in exploratory qualitative research [56]. Purposeful sampling was used to select organizations of different sizes and participants who had knowledge and experience within this specialized topic area. This was combined with convenience sampling, where participants were sought based on ease of accessibility to the researchers and their willingness to participate in the study.

Nine individuals were recruited from prior researcher contacts. Additional participants were recruited from among vendors at the RSA conference [14], a large industry IT security conference that also hosts an exhibition floor with security-focused vendors. A list of 54 potential organizations was compiled after in-person researcher contact on the exhibition floor. After the conference, we identified organizations that provided organizational diversity in our sample and that were accessible to the researchers. We emailed 17 of them to invite participation in the study. Eleven organizations agreed to participate. One additional organization was recruited based on the recommendation of a participant.

3.2 Interviews

We collected data via semi-structured interviews. Interviews were conducted by two of the researchers and ranged from 30 to 64 minutes, lasting on average 44 minutes. We conducted 21 interviews with 1-3 participants per interview, 29 participants total. Five organizations opted to have more than one participant in the interview: three organizations had three participants and two organizations had two participants. Face-to-face interviews were conducted if feasible. Otherwise, participants were given the choice of a phone or video conference interview. Five interviews were conducted face-to-face, 10 by phone, and six via video. Interviews were audio recorded and transcribed by a third-party transcription service.

After the first nine interviews, we performed a preliminary analysis and chose to make minor revisions to the interview

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

protocol in accordance with the qualitative research practice of theoretical sampling. Theoretical sampling involves adjusting data collection while the study is in-progress to better explore themes as they arise [13].

The interviews began with demographic questions about the organization (e.g., size, products) and the individual participants (e.g., role within the organization, professional background). Subsequently, participants were asked to describe their organizations' development and testing practices and associated challenges for their cryptographic products. Questions then transitioned into exploring cryptographic resources used by the organizations and how the participants thought those resources might be improved, if at all. The complete interview protocol is included in Appendix A.

3.3 Analysis

We utilized both deductive and inductive coding practices. Initially we constructed an *a priori* code list based on our research questions and literature in the field to provide direction in the analysis. As we performed multiple rounds of coding, we also identified emergent codes in the data. This iterative, recursive process helped us identify additional codes and categories as we worked with the data until we reached saturation [26, 69].

Five interviews (almost 24%) were first coded individually, then discussed as a group to develop a codebook. Although there is debate on the amount of text to collectively sample in qualitative research, the amount of text we group coded far exceeds the minimum of 10% often cited as standard practice [33]. We calculated intercoder reliability on this subset of the data using the ReCal3 software as a tool to help us refine our codes [21, 22]. For the five interviews, we reached an average Krippendorf's Alpha score of .70, with a high of .78, which is considered within the fair to good bounds for exploratory research having rich data with many codes and a larger number of coders [11, 15, 39].

Beyond the agreement metric, and in line with the views of many qualitative research methodologists, we thought it was important to focus on how and why disagreements in coding arose and the insights gained from discussions about these [5, 63]. These discussions better allow researchers to refine coding frames and pursue alternative interpretations of the data. During analysis, we found that each coder brought a unique perspective that contributed to a more complete picture of the data. For example, two of our coders more often identified high-level, nuanced codes about emotions and personal values, which may be due to their many years of working in human-focused contexts. These interpretations were often missed by the other coders who had more of a technology-focused background.

After coding of the initial subset of data, the remaining 16 interviews were coded by two coders each. Once each pair completed their coding, they had a discussion about the data to address areas of divergence about their use and application of the codes. This discussion resulted in the coders being able to understand each other's perspectives and come to a final coding determination. New codes that were identified during these discussions were added to the codebook, with previously coded interviews then re-examined to account for additions. The final codebook is included in Appendix B.

During the coding phase, we also engaged in writing analytic memos to capture thoughts about emerging themes [13]. For example, one memo captured thoughts on cryptography complexity. Once coding was complete, we reorganized and reassembled the data, created coding arrays, discussed patterns and categories, drew models, discussed relationships in codes and data, and began to move from codes to themes [59]. The team met regularly to discuss our emergent ideas and refine our interpretations. This process allowed for the abstraction of ideas and the development of overarching themes, such as how an organization's maturity and security culture drive formal development practices.

3.4 Limitations

Our study has several limitations. First, interviews are subject to self-report bias in which participants tend to underreport behaviors they think may be viewed as less desirable by the researchers, and over-report behaviors deemed to be desirable [16]. Given that the researchers who conducted the interviews represented an institution known for its security expertise, this bias may have influenced participant responses. We also note that the answers to some of the interview questions reflected participants' *perceptions* of the security level of their products and the security mindset of their organizations, which may or may not reflect reality.

Since there is no prior research into what is representative of the cryptographic development community, our sample is not characteristic of all types of organizations in this space. Although we did strive for diversity in organization size, with a smaller sample size common in qualitative research, we cannot definitively identify differences due to this variable.

4. DEMOGRAPHICS

Table 1 provides an overview of the organizations and participants in our study. To protect confidentiality, product types and participant roles are generalized.

The organizations represented in our study were of different sizes, with six being very large (10,000 or more employees), six large (10,00 - 99,99 employees), three medium (100 - 999 employees), three small (10 - 99 employees), and three very small/micro (1 - 9 employees) [24, 32]. All organizations developed a security product that uses cryptography (e.g. end user security software, hardware security module) or a non-security product that heavily relies upon cryptography to protect it (e.g. Internet of Things devices, storage devices, operating systems). Customers of these products ranged widely and included consumers, other parts of the organization, and organizations and businesses in multiple sectors such as government, technology, health, finance, automotive, and retail. Of the 15 organizations that discussed how long their companies had been implementing cryptography in their products, 12 had 10 or more years experience, with six of those having at least 20 years experience. The remaining three were startup companies that had been doing cryptographic development since their inception.

The 29 participants were a highly experienced group with several having made major contributions to the cryptography field. All participants had technical careers spanning 10 or more years, with several having been in the field for 30+ years. At least one individual from each of the interviews either currently worked on cryptography and security as a major component of their jobs (19 participants), or had

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

Table 1: Interview Demographics

ID	Org Size	Reg	Prod Type	Participant(s)
C01	VL	U.S.	HW	Lead crypto architect
C02	VL	U.S.	COM	Lead cryptographer
C03	VL	U.S.	HW, SW	Systems architect
C04	VL	U.S.	HW	Crypto design reviewer
C05	VL	U.S.	HW	Crypto architect
C06	VS	U.S.	SW	Systems analyst
C07	VS	U.S.	COM	Founder & researcher
C08	VS	U.S.	IOT	Founder & developer
C09	VL	U.S.	IOT	Researcher
C10	L	U.S.	SW	Founder & engineering lead
C11	L	U.S.	HW, SW	Product manager
C12	S	U.S.	SW	1) CTO 2) Marketing engineer 3) Business manager
C13	S	U.S.	SW	1) Chief Evangelist 2) Strategy Officer
C14	M	U.S.	SW	1) Marketing lead 2) Developer 3) Quality assurance
C15	L	U.S.	SW	Principal engineer
C16	L	U.S.	SW	CISO
C17	M	Eur	SW	1) CTO 2) Security engineer
C18	L	Aus	SW	Crypto engineer
C19	L	U.S.	SW	CTO
C20	S	U.S.	COM	1) Founder & architect 2) Compliance lead 3) Marketing director
C21	M	Eur	SW	Crypto specialist

Org Size: VL=Very Large, M=Medium, S=Small, VS=Very Small/Micro. **Reg** (Region/location of participant): U.S.=United States, Eur=Europe, Aus=Australia. **Prod (Product) Type:** HW=Hardware, SW=Software, COM=Communications Security, IOT=Internet of Things.

worked on cryptography extensively in the past (3 participants). The other participants were marketing or product leads, but all had a technical background.

Most of the participants had learned cryptography "on-the-job" as opposed to having formal training in the field. Five had an education in mathematics, but only two of those had studied cryptography as part of their formal study. Three had an engineering education, one had a physics degree, and the rest were educated in a computer-related discipline.

Four out of the 29 total interview participants had enhanced their knowledge through involvement in cryptographic standards groups. A cryptography architect commented on the value of his involvement in IEEE cryptographic standards early in his career: *"That's where I got to commune with cryptographers for a couple of years, me on the engineering side, and them on the crypto side... You end up learning things as a result of that process" (C05).*

5. RESULTS

The interviews revealed an organizational security mindset seldom seen in other cryptographic development studies. Our results suggest that the security mindsets had their roots in organizational attributes and culture that lay the foundation for the selection and use of cryptographic resources and rigorous development and testing practices.

For this section, we report counts of interviews throughout, joining group interview participants' answers to account for their organization. Due to the semi-structured nature of the interviews, the counts do not indicate quantitative results, but are reported to give weight to certain themes that were mentioned across interviews.

5.1 Security Mindset Characteristics

The interviews revealed organizational and personal characteristics that demonstrated a strong security mindset. These characteristics included professional maturity gained through experience, a deep understanding of the complexity of cryptography, and evidence of a strong security culture.

5.1.1 Emphasis on Experience and Maturity

Bruce Schneier said, "Only experience, and the intuition born of experience, can help the cryptographer design secure systems and find flaws in existing systems" [61]. The study participants expressed the importance of this experience as they repeatedly highlighted their own and their organizations' substantial maturity with respect to developing secure products and working with cryptography.

Overall, the organizations placed great value on hiring and retaining experienced technical staff. As previously mentioned, the majority of participants had substantial individual experience with cryptographic products. They also tended to work with other seasoned individuals. One participant described his team: *"We have a couple of the same core people on our test team who've been here for 25 years. They've gotten very good" (C01).* A startup company had only a few employees, but they all were veteran security software developers: *"Everyone we have has a lot of experience. I think the most junior person has a master's degree... and 10 and a half, 11 years of experience" (C07).*

Eight interviews noted the importance of experience when doing secure development, especially with cryptography. One participant remarked that secure products are ultimately dependent on *"the people that are designing it having the necessary knowledge and experience and understanding the whole picture, not just the little microscopic piece they're working on" (C01).* An interviewee with a long cryptographic background emphasized that there is no substitute for experience as he recounted a story of how his former company had to hire three less-qualified, full-time people to replace him in the work he had been doing on a part-time basis. A company founder remarked about the high level of technical maturity needed to properly deal with the complexity of cryptography: *"The level of education somebody needs to attain to be effective at doing crypto is relatively high. So it's not like I can put somebody who's fresh out of school on something and expect good results" (C10).*

5.1.2 Recognition of Cryptography Complexity

Based on their own experiences, participants and their organizations were keenly aware that, even though developing secure cryptographic implementations may appear to be easy, it is deceptively difficult. Despite proven algorithms being available, *"the algorithms are fairly involved,*

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

and they're difficult to understand" (C20). Yet understanding the algorithm is just the first step. As described by an IoT researcher, translating the algorithm correctly into a product is *"not a trivial implementation" (C09).* One design reviewer remarked about the pervasiveness of cryptographic design errors he encountered over the course of his career: *"I think I reviewed about 2,500 products... I can only remember eight that did not have a problem that either... they had to fix, or... they had to change their marketing claims" (C04).*

Our interviews also suggest that building cryptographic systems appears to be more of an art than a science, requiring a careful balance between security, performance, and usability. One participant described these tensions:

> *"Crypto algorithms are already very highly optimized ... It's like balancing a supertanker on a 40,000-foot high razor blade, and if you make one small change you destroy the performance. If you make it the other way, you just destroy the security." (C04)*

Because of the complexity, our participants recognized that design and implementation errors can be rampant and require rigorous review and testing to answer the misleadingly simple question *"How do you know it's right?" (C08).* However, assessing cryptographic products can be challenging, requiring knowledge and experience to construct good tests. A cryptographic development architect described challenges his organization had in the past when they lacked maturity in cryptographic implementation and testing:

> *"We actually implemented a new symmetric encryption algorithm, and it passed all the tests... and it turned out that they did the algorithm completely wrong. That was because... they wrote a test which said, 'Generate some random data, encrypt it with the algorithm, decrypt it, and see if you get back the original data.' Well, yeah, it got back the original data, but the encrypted data was incorrect. It just was symmetric, so it did the same wrong thing encrypting and decrypting." (C01)*

The organizations also understood that cryptography is just one of many interdependent product components, with all of them having to be properly implemented to ensure security. This sentiment was echoed by one participant who commented, *"For us, the design of the overall architecture that uses the crypto algorithms is almost as important as the correctness of the underlying algorithms themselves" (C01).*

5.1.3 Security Culture

Each organization in our study appeared to have a strong security culture that was interdependent on the maturity and experience of its employees. A security culture is a subculture of an organization in which security becomes a natural aspect in the daily activities of every employee [60]. For development organizations, this involves having dedicated security people, spending money on security, making security a company core value, and offering secure products. For the studied organizations, the culture included a commitment to address security and the perpetuation of a security mindset to others in the organization.

Commitment to Security: The organizations we studied thought that having good product security and strong cryptographic implementations was a *"core value" (C07)*, *"the key to quality" (C09)*, and essential to company identity. As an example, a Chief Information Security Officer (CISO) of a large company remarked, *"In our company, we are developing and selling security to our customers. So we care about, basically, all three sides of the sort of security triangle [confidentiality, integrity, availability] in what we do." (C15).* A security engineer talked about his company's belief that security must be an important consideration even when faced with competing tensions such as time-to-market: *"Since we do a [security product], everybody feels that we need to add security and good crypto at every step, so it's not a big issue to find the right balance" (C17).* Another participant commented on how his company demonstrates its commitment to secure cryptography: *"We have some fairly large teams which concern themselves with cryptography and secure design methodology. All engineers get training on secure design and we make it a big deal in the company" (C05).*

Security culture is not just internally-motivated. External motivators, like gaining a larger market share or customer requirements, often necessitate strong attention to security. One participant commented on how customer expectations fostered a security culture that drove rigorous testing processes within his company:

> *"We serve the kinds of customers who rely on the stuff to work reliably and properly from the get-go, when they buy it. So it's not like... 'Maybe we'll update something later, if we find some problems.' That's not our philosophy, and that's not what our kind of customers expect... Part of that is also company culture." (C01)*

Although security culture is often thought of as a "top-down" phenomenon, it must be accepted by and acted upon at all levels. One participant, a CISO for a large company, commented on the importance of the security culture being pervasive throughout an organization:

> *"If there's senior executive support for a strong security program,... that helps tremendously. At the same time, if there is still a very strong feeling amongst a large number of developers that security, cryptography, and everything that's related to that is really a nuisance that should get out of the way and just to prevent them from writing more interesting features, it's definitely a concern." (C16)*

The interviews showed evidence that our participants are critical to the "bottom-up" support of organizational security culture. They serve as security champions and self-appointed educators, leading by example and projecting their values, personal philosophies, and commitment to security on the rest of the organization. Two of the participants explicitly embraced this role as a personal mission when they referred to themselves as "security evangelists." Another expressed his feeling of personal accountability to enact security in the products he supported: *"It's essentially a mark of my success or not, that I'm measured against, of whether those things remain secure or hacked" (C05).*

Perpetuating a Security Mindset: Just as the employees influence security culture, so does the culture influence employees by perpetuating a security mindset. Part of this perpetuation involves expert employees mentoring and supporting less-experienced personnel in their learning of secure programming methods and specialized security topics such

as cryptography, as discussed in ten interviews.

The interviews suggest that providing an opportunity for individuals to gain hands-on experience with real products is important in understanding the issues involved in developing a cryptographic product. However, given the distinct possibility that a novice will make errors in the implementation, precautions must be taken. Two participants suggested that mock training exercises conducted on a separate testing infrastructure may be valuable initial steps. Others discussed mentoring and peer review activities within their organizations. For example, one organization enacted *"parent programming" (C14)* for any code that uses cryptography. Another had a formal peer review process:

> *"We review everyone's code... When I write something down and it has a flaw in it, I'm told about it, which is good... I think what we do is we take smart people who care about doing good work, and we foster an environment where they're not afraid to receive constructive criticism, and they're not intimidated away from giving it." (C15)*

The influence of an organization's security mindset does not necessarily end when an individual leaves that organization. As evidenced by three participants who left large companies to start their own small businesses, there seemed to be a transfer of security culture and practices from previous employers. A small company founder described this transfer: *"I think some of it could be just kind of from my career background, what I learned were the best practices... I think we've just kind of learned them in the beginning and kind of kept that culture" (C08).*

Size Doesn't Matter: We found no significant differences in perceived security culture or overall security practices based on size. Obviously, larger organizations had more resources to dedicate to security and cryptographic development. However, smaller organizations understood that vulnerabilities in their products could do great harm to the company's reputation, and so were committed to security and made thoughtful decisions about how they developed and tested, even if on a smaller scale. For example, the founder of a micro business noted:

> *"Being a small company, we're trying to also gain credibility. And we don't want to just claim that we have the fastest [crypto implementation] in the world, we also want to make sure that it is built safely and validated... [W]e cannot afford for this thing not to work properly." (C07)*

5.2 Selection of Resources

Because of security mindsets, participants revealed a proclivity towards careful selection of resources to help them attain their goal of secure cryptographic implementations. In this section, we describe considerations taken when choosing and evaluating cryptographic resources.

5.2.1 Standards

In line with the popular quote "The nice thing about standards is that you have so many to choose from" by Andrew S. Tanenbaum [67], the interviews revealed that all the organizations used some type of cryptographic standards from organizations such as IEEE, ISO, American National Standards Institute (ANSI) [3], Internet Engineering Task Force (IETF) [37], and Payment Card Industry (PCI) [57]. All but one described using NIST standards or guidance documents, most commonly FIPS 140-2. The participants and their organizations were knowledgeable enough to understand and evaluate the appropriateness of cryptographic standards. However, not all organizations have the need to directly read the standards. For those that do, this requires maturity in the field given the standards' complexity. A Chief Technology Officer (CTO) at a small company expressed this challenge: *"A lot of standards are notoriously difficult to read. Unless you're an expert in the field, a lot of them don't make sense" (C12).* In another interview, a principal engineer commented on the difficulty in translating standards to products:

> *"I can tell you from my personal experience understanding the fundamentals of these things, still the standards were a challenge to use because they were very divorced from the implementation day-to-day details that I encounter while I'm trying to plug all the pieces together." (C15)*

Despite the complexity, when selected carefully and implemented correctly, standards were seen as beneficial for several reasons noted by our participants. Participants in eight out of 21 interviews commented that the community review of standards results in a more correct and secure solution. A CISO at a large software as a service company reflected on the value of public scrutiny: *"By relying on other standards that [were] vetted by multiple parties, I have much higher assurance that the underlying cryptography and design [are] done in an appropriate way" (C16).* A director of product management concurred with this: *"So the standards, because it's out there and everybody's looking at it and testing it, we depend on that as kind of a layer of security" (C14).*

Included in community review is the transparency of the standards process. One participant illustrated this observation using the popular AES standard as an example:

> *"It gave us a lot of comfort knowing how AES evolved ... being able to see all the steps, having that all happen out in the open and why and how it happened. Very helpful to us making the decision for what we're going to use and why we're going to use it." (C10)*

Participants in nine out of 21 interviews commented that the use of standards eliminates some of the difficulty of cryptographic development and testing by providing an authoritative foundation. The founder of a software company said his organization relies heavily on standards because *"inventing it on your own is just a different level of complexity that we knew enough to know we did not want to be involved in" (C10).* Standards also add confidence that a product will be *"interoperable with our customers, with our partners, even with our competitors" (C16).*

Finally, participants noted that standards-based products may elicit customer confidence. The director of product line management at a large credential management company spoke of the importance of customer trust in gaining market share, saying, *"If the standard is mature, then it means our product's going to be more easily accepted by customers" (C11).*

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

Standards meet the needs of most companies we interviewed; however, there are cases in which standards fail to address a specific need. In these situations, organizations may extend or modify standards. These extensions were viewed as adding rigor and security in addition to functionality, making the cryptographic implementations *"really ahead of any industry standard practices" (C05).*

Interestingly, three participants commented on distrust of government standards because of allegations that a U.S. government agency purposely weakened cryptographic algorithms [28]. For example, although one organization made extensive use of standards, they took special measures to exclude aspects of a government standard they felt were questionable and exercised extra rigor in their testing processes to account for potential vulnerabilities. In an extreme case, a consulting company's observation of customer distrust of government standards and their own frustration with the complexity of those led them to develop a new cryptographic primitive: *"I think it [the distrust] comes from . . . news reports or exposés that say this standard may not be as secure as we think. . . There is a lot of doubt out there . . . That's why there has to be additional options and alternatives" (C06).*

5.2.2 Certifications

Seventeen out of 21 organizations obtained at least one formal certification that the implementation of cryptographic algorithms in their products met standards specifications. Three additional organizations developed and tested to certification criteria without undergoing full certification. Eighteen organizations referenced FIPS 140-2 certification [49], five Common Criteria [41], three the Payment Card Industry Data Security Standard (PCI-DSS) [57] validation program, one the Underwriters Laboratories (UL) certification [70], and others pursued country-specific certifications.

The perception of the benefit provided by adhering to certification requirements was mixed among our participants. Among those who obtain certifications, only five organizations expressed that certifications establish additional confidence through independent testing. One of these remarked, *"You have a lot of assurance that everything's going to be tested and get that nice, kind of warm and fuzzy" (C08).* Six organizations noted that, even though they do not undergo the formal FIPS 140-2 certification process, they build to and test against the certification specifications to gain added assurance. A participant from a key and identity management software company stated, *"as a small company, I think it is actually extra important to make sure that we go through this battery of tests just to in a way reassure the people we're talking to that this is a robust product" (C12).*

For some participants, certifications are perceived as being more useful for meeting customer expectations than for bolstering security. Organizations most often obtain certifications because these are requirements of their customers in certain sectors (e.g. government, financial): *"for some areas, if you don't get the check-mark you don't get to play" (C11).*

Unfortunately, as noted in 12 of 21 interviews, certifications can be expensive in time and resources, making them prohibitive for smaller companies, especially those with products that run on multiple platforms and have frequent version updates. However, our interviews suggest that confidence in the cryptographic implementations may not be dependent on any certification, but rather on the rigorous development and testing practices these organizations undertake. The founder of a small company commented that FIPS 140-2 certification was too costly for them to pursue, but was confident that his product met the certification requirements: *"It's a place where we've done enough testing ourselves. I know it's fine. I know we would pass" (C10).* Another participant, whose company did undergo certifications, felt that the certifications provided little assurance beyond what was provided by their own internal processes:

> *"We always design and test ourselves to have confidence that [the product] meets all those requirements before we release it to the lab for their testing. So when they come back and say, 'It passed this,' we say, 'Well, okay, we expected that. Thank you.' So the surprise is if something fails, that we expected to pass. That doesn't happen very often." (C01)*

Four of our participants also remarked that certifications may not be a robust contributor to the security of a product. They expressed strong sentiments that certifications, especially FIPS 140-2, are more of a "checkbox" for customers *"without any additional benefit of security" (C02).* A CTO and long-time cryptographer added, *"FIPS 140 is. . . not focused on how to use crypto securely. It's focused on how to safely provide crypto functionality" (C19).*

In addition, seven out of 21 organizations commented that maintaining FIPS certification may, in some cases, weaken security by discouraging updates throughout the lifecycle of the product. Once a product undergoes a significant update (for example, fixing a security vulnerability), it may lose its certification. Organizations are then put in a difficult position: *"this ability to address vulnerabilities and to patch validated code is a real problem. It sends the wrong message if you do what you should do, which is patch it and live without [the certification]" (C19).*

5.2.3 Third-Party Implementations

The complexity of implementing algorithms from scratch and the expertise required to write those compelled two thirds of the organizations to follow industry best practices by not writing their own cryptographic code and instead using third-party cryptographic implementations, for example open source libraries such as OpenSSL or built-in operating system APIs. One participant used an analogy to explain why his organization used these resources: *"Not every person should be performing brain surgery on another person. I also don't believe every software engineer should really go write crypto code" (C07).*

The organizations selected these third-party implementations based on several factors. First, four mentioned an implicit trust of the resource based on the reputation of the vendor or general community acceptance. In describing why his organization chose a set of cryptographic libraries, one participant said, *"You pretty much trust those libraries because they are widely used, and you can run test vectors against them easily" (C20).* A third of the organizations said that they have more confidence in formally vetted, certified implementations. A participant from a small company that works on IoT cryptography commented, *"If a vendor has submitted a library through FIPS 140-2 certification, versus a code that was up on GitHub for example,. . . I would*

be more inclined to trust the one that has gone through the FIPS validation" (C08).

Despite benefits of using third-party implementations, the organizations respected that the use of these external resources can still be difficult for less-knowledgeable individuals. A developer provided an example:

> *"[Crypto libraries] in general don't provide enough to be able to use them correctly out of the box... But there's many out there that think that they can just use AES, and I included it and I'm using it. But I'm not using it correctly, and then I'm leaving myself open to attack." (C14)*

This point illustrates that there is an important distinction between a cryptographic algorithm being certified or deemed "secure" and the proper use of the algorithm by developers. Similarly, third-party libraries have faced their share of security vulnerabilities due to implementation errors of otherwise sound cryptographic algorithms. Some of these vulnerabilities have had far-reaching impacts, for example the Heartbleed vulnerability in OpenSSL [71]. A security architect commented that third-party implementations are *"much more likely to contain implementation bugs and vulnerabilities" (C20).* Therefore, some organizations attempted to vet third-party implementations by doing their own vulnerability checks. One participant noted that his organization monitors the vulnerability databases for security issues with the libraries they use. Another commented on the extra security checks his company performs: *"We validate outside third-party libraries and software as they come in and confirm that they are bug-free and up to the latest standard or perform the right risk assessments" (C16).*

Finally, organizations may augment third-party implementations with their own internally developed modifications to avoid potential errors or address gaps in the resources. For example, to prevent developers from making errors while using the Windows CryptoAPI, a vulnerability assessment software company had *"written libraries on top of that to present a prettier facade in front of it because it's a fairly difficult library to use the way it's delivered" (C15).*

5.2.4 Academic and Research Resources

Our interviews revealed differing views on the value of academic research resources. Participants in three out of 21 interviews said that they have referenced academic resources (e.g., attended academic conferences or read (attack) research papers) to either better understand cryptography or to keep up with advances in the field. However, other participants passionately voiced their lack of confidence in the relevance of cryptographic research to their organizations' real-world industry challenges. One participant commented, *"People out in academia are famous for claiming there are holes in this kind of stuff where they don't actually exist, because they don't configure things according to the recommendations" (C01).* A cryptography architect asserted that his company's testing methods for cryptographic implementations were more state-of-the-art than those described in the academic world: *"No, we don't reference academic papers. They're not where we are in understanding the test problem... So there's a six-year gap between the... methods that we developed being identified in academia" (C05).*

5.3 Development and Testing Rigor

Our interviews revealed that the organizations translated their commitment to security and their expertise into rigorous development and testing practices. Interestingly, when participants spoke about how they test the cryptographic components, they saw secure software development as the foundation to providing cryptographic functionality.

5.3.1 Formal Processes

Of our 21 interviews, 20 reported that their companies employed formal development and testing strategies (those that are structured and standardized within the company) to ensure that their products, including the cryptographic components, were secure, while one participant said that they contracted developers to do that for them.

The development and testing practices were often the result of an evolutionary process spanning many years, as noted by 10 interviews. A director of quality assurance remarked, *"Part of [the role of] our test lead is now to verify that we're at the appropriate levels from a security standpoint... so it's a big focus for us now, whereas in the past, it was kind of a side item" (C14).* A company founder described his organization's introduction of more robust techniques over time:

> *"And it was an evolution, honestly. It started to where we didn't have hardly anything and new tools came on the market, as well as we had more time to focus on it. That allowed us to kind of improve code incrementally over time." (C10)*

Twelve interviews revealed a strong security mindset when they described developing and testing their products' cryptographic components based on risk. They would carefully build threat models to protect against strong adversaries, perform penetration testing, and would monitor current vulnerabilities to ensure their products were secure against those. A cryptographic design reviewer commented on this risk-based approach: *"All we can do is try to build good adversary models and then try to determine if our systems can stand up to those adversaries" (C04).*

Another discussed how his company's practices ensured that security had been considered throughout development:

> *"We have architects that do security reviews, that do threat modeling. And it's not just about the crypto, but more in general, how do you use the product. Who gets to do what? What are the risks? How do we mitigate those risks?... And one of the items for the engineering gate release is making sure... we mitigated anything that needed to be mitigated." (C11)*

Generally, the organizations' development and testing processes adhered to the principles in Figure 1. First, security specifications, architectures, and designs would be developed and reviewed. These would then be programmed against. Internal or external code reviews would be subsequently be conducted. During testing, tests would be run using internally developed test vectors or those provided with cryptographic resources. Static analysis tools and testing tools were also generally used. This development and testing process would be iterated whenever functionality changed, and performed in an expedited fashion if updates were being made due to a discovered security vulnerability.

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

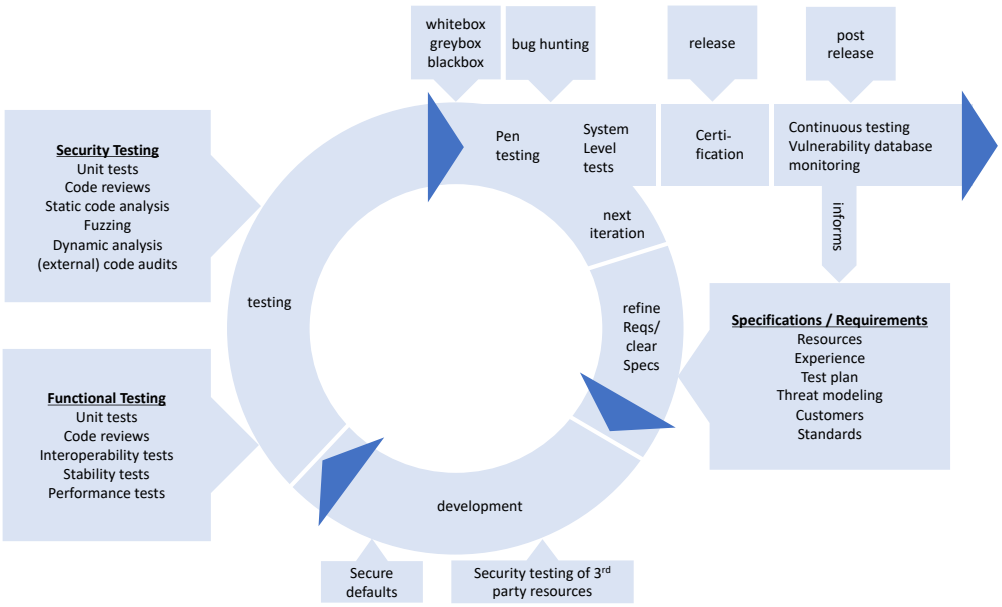


Figure 1: Security Development Lifecycle [34] adapted to include development and testing strategies as extracted from the interviews. Not all processes apply to all interviewed organizations.

5.3.2 *Development*

As mentioned above, the development phase for the organizations followed typical, commonly accepted development practices such as requirements and risk analysis and programming. We specifically highlight two practices that were mentioned most often in the interviews.

Participants in nine interviews spoke about design reviews, which were critical for finding potential errors in cryptographic components early in the process. Those that do cryptographic review must be highly skilled and able to piece together others' thought processes:

> *"A lot of the review is really just archeology. It's delving down into what they're producing, trying to understand both the explicit and the implicit assumptions, and then identifying where there are conflicts that lead to attacks." (C04)*

Nine organizations also mentioned the importance of doing code reviews to look for security and functional errors and vulnerabilities. A participant from a security software company remarked, *"We have a mandatory and systematic code review. Each line of code and each comment of code needs to be reviewed by usually at least two peers" (C17).*

5.3.3 *Testing*

Figure 2 shows the types of testing mentioned in the interviews. In 16 interviews, automated testing was discussed, often as being integrated with manual testing. The CTO of a company that produces security software discussed this integration of automated and manual testing: *"We have automatic tests, unit tests, integration tests, functional tests . . . , code analysis. . . . We have additional, manual tests being done. . . on top of the automated tests" (C17).*

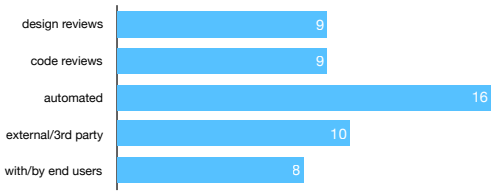


Figure 2: Development and testing practices explicitly mentioned for secure cryptography.

Ten interviews mentioned the use of third-party testing to improve the security of their cryptographic products. For example, organizations used bug-hunting services, or external testing such as blackbox, greybox, or whitebox penetration tests to increase the chances that bugs would be found in a controlled environment, and prior to product release. One participant described the benefit of his company using a bug-finding platform:

> *"You have some complete geniuses there that have found things that we never would have found. . . I feel like you're way more trustworthy if you are actually upfront about this stuff and you are actively soliciting people to attack you and paying them for their effort. You get a much higher confidence that some random person attacking won't just find something easy." (C10)*

Third-party review can also be used to gain trust with customers as expressed by a product manager: *"When customers ask us. . . 'Can you prove to me that it's done securely?' we can point to another organization that's independent to show" (C11).*

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

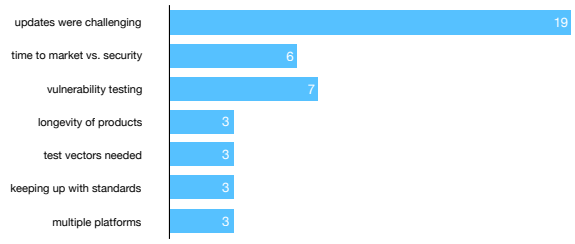


Figure 3: Challenges in development and testing.

End-user testing was not deemed as important for some organizations as they mostly develop products that become components in other products. However, this type of testing is critical for those producing products that will be directly used by businesses and consumers, and was discussed in eight interviews. These interviews mentioned formal beta-testing, continuous feedback, employing a user testing service, or recruiting convenience samples to test the product. One participant described the importance of user testing for his organization's consumer security software:

> *"We'd bring in our friends and family and sit them down and watch them. And it was eye-opening. It caused us to change a lot of what we did because, essentially, they didn't get the concept... We also used usertesting.com, which has been very great. You can show 10 people something and you have a pretty good idea." (C10)*

Another company takes advantage of beta testing to identify potential issues in their product: *"We have a very extensive beta program, and we have a very active customer base... so we have no lack of feedback" (C15).*

5.3.4 Challenges

All 21 of our interviews mentioned challenges to development and testing (Figure 3). We focus here on challenges directly related to cryptography, excluding challenges that have already been discussed, e.g., cryptographic complexity.

One challenge mentioned in six interviews was the tension between getting a product to market and taking the time to do robust security development and testing. For example, a participant from a company that spends years securely developing its cryptographic hardware modules observed that in most of the hardware/software industry, *"The design focus is simply not to do a solid job which will last a long time cryptographically.... They cannot care because they have to get the products out in a timely fashion" (C05).*

Testing for vulnerabilities in cryptographic implementations was another challenge mentioned in seven interviews, with three expressing concern for adequate testing of side-channel attacks. A participant who integrates cryptography into IoT devices said, *"especially in the embedded world, what the test vectors don't address I think is side-channel attacks, [which] could be really detrimental to the embedded device" (C08).*

Another challenge, as mentioned in three interviews, was the longevity of cryptographic products in customer spaces. An IoT researcher commented, *"Many devices are going to be deployed for 20 years. So, maybe the crypto in 20 years is no longer secure" (C09).* Another participant expressed the difficulty of having to maintain legacy cryptographic algorithms in a product: *"Old things become weak and you shouldn't use them anymore, and you need to add new ones ... However, customers are not so cooperative... It may take 10 years before everybody stops using something" (C01).*

Four interviews discussed challenges in having to troubleshoot or update third-party cryptographic implementations when vulnerabilities or errors are found. A principal engineer at a security software company described, *"when we're using any third-party library...and you happen to do something, and it fails, it's really hard to figure out what went wrong" (C15).*

Other cryptography-related challenges included the need for more test vectors (3 interviews), keeping up with changes in cryptographic standards (3), and having to use cryptographic libraries on multiple platforms (3).

In spite of these challenges, organizations in our study reported confidence in their processes and the resulting security of their cryptographic products (16 interviews). For example, a senior systems analyst at a small company described his confidence in the cryptographic algorithm they had developed: *"I don't make any bold claims, but at the same time, we looked at our encryption algorithm and we considered it quantum-proof" (C06).* In another interview, a company founder stated, *"I think we are definitely in the higher echelon for going above and beyond" (C10).*

6. IMPLICATIONS

6.1 Expanding Research Contexts

Our results suggested that the organizations believe they have a mature workforce, appreciate the complexity of crypto, possess a strong security culture, effectively use cryptographic resources, and practice secure development. However, the various studies mentioned in Related Work (e.g., [1, 2, 17–19, 42, 48]) indicate that there are many poor cryptographic implementations "in the wild," and developers typically lack a fundamental understanding of cryptography.

Where, then, is the disconnect between our findings and past research? It is possible that our self-report data are merely perceptions and do not accurately reflect the security mindsets of the organizations. Participants may be overconfident or may have overstated their organizations' security practices because of observer bias. We also recognize the value of future research to verify organizational claims by examining vulnerability databases, for example Common Vulnerabilities and Exposures (CVE) [68], to enumerate security issues in their products. Additionally, knowledge of an organization's development maturity level (e.g. Capability Maturity Model [12]) could be used as a comparison point.

Alternatively, this population is likely quite distinctive from previously studied developer populations and contexts, which may explain some of the differences in our findings. First, our participants exhibited more maturity in security and cryptography, with all having more than 10 years of security development experience. Second, organizational culture and constructs were an important driver of security mindsets within our study, while previous studies often involved

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

independent application developers, many of whom were not full-time developers or had not received formal education or organizational training in programming, cryptography or security. Third, there was a marked difference in the types of cryptographic resources used. Other studies (e.g., [2]) indicated that developers are reliant on information gleaned from search engines and Stack Overflow, which was only mentioned once in our interviews. Instead, the organizations in our study turned to more authoritative resources such as cryptographic standards and certification specifications. Finally, many of the studies that identified cryptographic vulnerabilities examined mobile apps, presumably because these were easy to obtain from public application stores, and open source projects. Our participants were developing more complex, expensive software and hardware products. Lack of security in these products had greater consequences, with the potential of harming the company's reputation or resulting in loss of sales.

These differences suggest that perhaps the security research community is not capturing a comprehensive picture of the cryptographic development environment. This demonstrates the need for the research community to diversify their study populations and contexts, and consider mechanisms to bridge the gap between more security-mature and less-skilled developers who implement cryptography in their products.

6.2 Support for Other Populations

The evidence of the criticality of organizational security culture and collaboration during development and testing raises the question of whether it is even possible for "solo" developers, such as the application developers with little cryptography experience or peer support represented in previous studies, to be truly successful in this area. How then can the research community explore ways to facilitate the transfer of strong security practices observed in some organizations to others with less support and experience?

As previously stated, unlike the population in our study, many developers sampled in past studies rely on online communities such as Stack Overflow [65] when implementing cryptography. But there is little evidence that these communities provide the level of support necessary for secure development. Future research may involve further assessing the value of current online communities for cryptographic development and exploring alternatives as means by which security mindsets can be created and perpetuated.

The findings also reveal that mentoring and peer review are critical to perpetuating security mindsets within organizations. Past studies have sought to understand the effectiveness of software development mentoring, peer review, and pair programming (e.g., [7, 47, 74]). However, more work needs to be done to determine whether mentoring for cryptographic development requires different tactics and how to best support this outside of organizational constructs.

6.3 Cryptographic Resource Usability

Whereas the bulk of responsibility in producing secure cryptographic products lies with the organizations themselves, our results imply that cryptographic resource providers can also do more to contribute to developer confidence. The most mentioned complaint our participants had with standards and certification guidance was the complexity of the language. This underlies a need for standards organizations to work towards a common language between cryptography experts who write the standards and developers and engineers who use them. Although standards documents may not be the appropriate place for large amounts of explanatory text, supplementary guidance that contains more instruction, cautions against common errors, and provides example implementations may prove to be valuable. Just as research has been done on language for security alerts and warnings (e.g. [10, 66]), it would also be helpful for researchers to explore the efficacy of language that explains cryptography concepts to non-experts.

Additionally, developers could benefit from more explanations of motivation, in other words, the "why" behind cryptographic choices. One participant echoed this recommendation as an important way to move beyond the "checkbox" mentality of standards: *"So you're thinking big picture, 'I'm doing this for this a reason,' because otherwise, you just get in the cookbook approach of, 'Do I meet this? Yes, yes, yes. Check, check, check' " (C19).*

Of course, many developers have no need to look at the standards directly since they use third-party implementations. However, third-party implementations may also be difficult to interpret and use securely if one lacks basic knowledge of cryptography. Our study results support past research calling for increased usability of cryptographic APIs. Similar to the work of Montandon et al. that proposed a new platform for providing API code examples [46], we also recommend investigating new approaches to community vetting of sample code since developers often copy flawed code snippets from forums such as Stack Overflow [2].

Notably, the participants spoke of a security problem with FIPS 140-2: updating certified software for security would break its certification, so companies relying on the certification had to decide between withholding an update or having to undergo recertification. This insight into an instance where reliance on a certification can decrease security underlines the need to closely and continuously involve cryptographic experts in the certification process. Their experience as users of the certification can help evolve the process and shape it to be more resilient, usable, and secure.

7. CONCLUSION

Our study offers new insight into the cryptographic development and testing practices of a previously unstudied population of organizations and participants who were highly experienced in cryptographic development. Our results suggest that organizational security mindsets are based on maturity and a strong security culture, which in turn guide selection of cryptographic resources and inform rigorous development and testing practices. Based on these observations, we see opportunities for development organizations, cryptographic resource providers, and security researchers to facilitate an environment conducive to building the expertise required to correctly and securely implement cryptography.

Disclaimer

Certain commercial companies or products are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the companies or products identified are necessarily the

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

best available for the purpose.

Acknowledgements

The authors would like to thank the following individuals who offered insightful comments that helped improve the quality of this paper: the anonymous reviewers, Curt Barker, Sascha Fahl, Simson Garfinkel, Michelle Mazurek, Matthew Smith, Brian Stanton, and Jeff Voas.

8. REFERENCES

[1] Y. Acar, M. Backes, S. Fahl, S. Garfinkel, D. Kim, M. Mazurek, and C. Stransky. Comparing the usability of cryptographic APIs. In *Proceedings of the 38th IEEE Symposium on Security and Privacy*, 2017.

[2] Y. Acar, M. Backes, S. Fahl, D. Kim, M. L. Mazurek, and C. Stransky. You get where you're looking for: The impact of information sources on code security. In *Proceedings of the 37th IEEE Symposium on Security and Privacy*, pages 289–305, May 2016.

[3] American National Standards Institute. ANSI. `https://www.ansi.org/`, 2018.

[4] S. Arzt, S. Nadi, K. Ali, E. Bodden, S. Erdweg, and M. Mezini. Towards secure integration of cryptographic software. In *Proceedings of the 2015 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (Onward! '15)*, pages 1–13, 2015.

[5] R. S. Barbour. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal*, 322(7294):1115–1117, 2001.

[6] C. A. Barry, N. Britten, N. Barber, C. Bradley, and F. Stevenson. Using reflexivity to optimize teamwork in qualitative research. *Qualitative Health Research*, 9(1):26–44, 1999.

[7] A. Begel and B. Simon. Novice software developers, all over again. In *Proceedings of the Fourth International Workshop on Computing Education Research*, pages 3–14, 2008.

[8] D. J. Bernstein, T. Lange, and P. Schwabe. The security impact of a new cryptographic library. In *Proceedings of the International Conference on Cryptology and Information Security (LatinCrypt '12)*, pages 159–176, 2012.

[9] E. Bonver and M. Cohen. Developing and retaining a security testing mindset. *IEEE Security & Privacy*, 6(5), 2008.

[10] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, 9(2):18–26, 2011.

[11] H. Brenner and U. Kliebsch. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology*, pages 199–202, Mar. 1996.

[12] CMMI Institute. What is capability maturity model integration (CMMI)? `http://cmmiinstitute.com/capability-maturity-model-integration`, 2018.

[13] J. Corbin and A. Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Thousand Oaks, CA, 4th edition, 2015.

[14] Dell Incorporated. RSA Conference: Where the world talks security. `https://www.rsaconference.com/`, 2018.

[15] K. DeSwert. Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. Technical report, Center for Politics and Communication, 2012.

[16] S. I. Donaldson and E. J. Grant-Vallone. Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17(2):245–260, 2002.

[17] T. Duong and J. Rizzo. Cryptography in the web: The case of cryptographic design flaws in ASP.NET. In *Proceedings of the 31st IEEE Symposium on Security and Privacy*, pages 481–489, May 2011.

[18] M. Egele, D. Brumley, Y. Fratantonio, and C. Kruegel. An empirical study of cryptographic misuse in Android applications. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*, pages 73–84, 2013.

[19] S. Fahl, M. Harbach, T. Muders, L. Baumgartner, B. Freisleben, and M. Smith. Why Eve and Mallory love Android: An analysis of Android SSL (in)security. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS '12)*, pages 50–61, 2012.

[20] C. Forler, S. Lucks, and J. Wenzel. Designing the API for a cryptographic library: A misuse-resistant application programming interface. In *Proceedings of the 17th Ada-Europe International Conference on Reliable Software Technologies (Ada-Europe '12)*, pages 75–88, 2012.

[21] D. Freelon. Recal3: Reliability for 3+ coders. `http://dfreelon.org/utils/recalfront/recal3/`.

[22] D. G. Freelon. ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1):20–33, 2010.

[23] R. Garcia, J. Thorpe, and M. Martin. Crypto-Assistant: Towards facilitating developer's encryption of sensitive data. In *Proceedings of the 12th Annual International Conference on Privacy, Security and Trust (PST '14)*, pages 342–346, 2014.

[24] Gartner. IT glossary: Small and midsize business (SMB). `http://www.gartner.com/it-glossary/smbs-small-and-midsize-businesses/`, 2017.

[25] M. Georgiev, S. Iyengar, S. Jana, R. Anubhai, D. Boneh, and V. Shmatikov. The most dangerous code in the world: validating SSL certificates in non-browser software. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, pages 38–49, Oct. 2012.

[26] B. G. Glaser and A. L. Strauss. *The Discovery of Grounded theory: Strategies for Qualitative Research*. Transaction Publishers, 2009.

[27] M. Green and M. Smith. Developers are not the enemy! The need for usable security APIs. *IEEE Security & Privacy*, 14(5):40–46, Sept. 2016.

[28] L. Greenemeier. NSA efforts to evade encryption technology damaged U.S. cryptography standard. `https://www.scientificamerican.com/article/nsa-nist-encryption-scandal/`, Sept. 2013.

[29] G. Guest, A. Bunce, and L. Johnson. How many

interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1):59–82, 2006.

[30] P. Gutmann. Lessons learned in implementing and deploying crypto software. In *Proceedings of the 2002 Usenix Security Symposium*, pages 315–325, 2002.

[31] J. M. Haney, S. L. Garfinkel, and M. F. Theofanos. Organizational practices in cryptographic development and testing. In *Proceedings of the 5th IEEE Conference on Communications and Network Security (CNS '17)*, Oct. 2017.

[32] B. Headd. The role of microbusinesses in the economy. `https://www.sba.gov/sites/default/files/`, Feb. 2015.

[33] R. Hodson. *Analyzing Documentary Accounts (No. 128)*. Sage Publications, 1999.

[34] M. Howard and S. Lipner. *The security development lifecycle Vol. 8*. Microsoft Press, Redmond, WA, 2006.

[35] Institute of Electrical and Electronics Engineers. IEEE Standards Association. `http://standards.ieee.org/`, 2018.

[36] International Organization for Standardization. Standards. `https://www.iso.org/standards.html`, 2018.

[37] Internet Engineering Task Force. IETF. `https://www.ietf.org/`, 2018.

[38] S. L. Kanniah and M. N. Mahrin. A review on factors influencing implementation of secure software development practices. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 10(8):3022, 2016.

[39] K. Krippendorff. *Content Analysis: An Introduction to its Methodology*. Sage, 2004.

[40] D. Lazar, H. Chen, X. Wang, and N. Zeldovich. Why does cryptographic software fail?: A case study and open problems. In *Proceedings of the 5th Asia-Pacific Workshop on Systems (APSys '14)*, pages 7:1–7:7, 2014.

[41] D. Leaman. National Institute of Standards and Technology: NVLAP Common Criteria testing. NISTHB 150-20. `https://www.nist.gov/sites/default/files/documents/nvlap/NIST-HB-150-20-2014.pdf`, 2014.

[42] Y. Li, Y. Zhang, J. Li, and D. Gu. iCryptoTracer: Dynamic analysis on misuse of cryptography functions in iOS applications. In *Proceedings of the International Conference on Network and System Security*, pages 349–362, Oct. 2014.

[43] G. McGraw. Software security. *IEEE Security & Privacy*, 2(2):80–83, 2004.

[44] C. G. Menk. System security engineering capability maturity model and evaluations: Partners within the assurance framework. `https://csrc.nist.gov/csrc/media/publications/conference-paper/1996/10/22/proceedings-of-the-19th-nissc-1996/documents/paper010/cmmtpep.pdf`, 1996.

[45] S. Merriam and E. Tisdell. *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons, San Francisco, CA, 4 edition, 2016.

[46] J. E. Montandon, H. Borges, D. Felix, and M. T. Valente. Documenting APIs with examples: Lessons learned with the APIMiner platform. In *Proceedings of the 20th IEEE Working Conference on Reverse Engineering (WCRE)*, pages 401–408, 2013.

[47] M. M. Müller. Two controlled experiments concerning the comparison of pair programming to peer review. *Journal of Systems and Software*, 78(2):166–179, 2005.

[48] S. Nadi, S. Krüger, M. Mezini, and E. Bodden. Jumping through hoops: Why do Java developers struggle with cryptography APIs? In *Proceedings of the 38th International Conference on Software Engineering (ICSE '16)*, pages 935–946, 2016.

[49] National Institute of Standards and Technology. FIPS Pub 140-2: Security requirements for cryptographic modules. `http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf`, 2001.

[50] National Institute of Standards and Technology. Cryptographic module validation program. `https://csrc.nist.gov/projects/cryptographic-module-validation-program`, 2016.

[51] National Institute of Standards and Technology. Cryptographic algorithm validation program (CAVP). `="http://csrc.nist.gov/groups/STM/cavp/`, 2017.

[52] National Institute of Standards and Technology. Cryptographic standards and guidelines. `https://csrc.nist.gov/Projects/Cryptographic-Standards-and-Guidelines`, 2018.

[53] P. Nguyen. Can we trust cryptographic software? Cryptographic flaws in GNU privacy guard v1. 2.3. In *Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques*, pages 555–570, 2004.

[54] Open Web Application Security Project. OWASP secure software development lifecycle project. `https://www.owasp.org`, 2017.

[55] OpenSSL Software Foundation. OpenSSL cryptography and SSL/TLS toolkit. `https://www.openssl.org/`, 2018.

[56] M. Q. Patton. *Qualitative research*. John Wiley & Sons, San Francisco, CA, 2005.

[57] PCI Security Standards Council. PCI Security. `https://www.pcisecuritystandards.org/pci_security/`, 2018.

[58] B. Potter and G. McGraw. Software security testing. *IEEE Security & Privacy*, 2(5):81–85, 2004.

[59] J. Saldaña. *The Coding Manual for Qualitative Researchers*. Sage, 3 edition, 2015.

[60] T. Schlienger and S. Teufel. Information security culture-From analysis to change. *South African Computer Journal*, (31):46–52, 2003.

[61] B. Schneier. Why cryptography is harder than it looks. `https://www.schneier.com/essays/archives/1997/01/why_cryptography_is.html`, 1997.

[62] B. Schneier. The security mindset. `https://www.schneier.com/blog/archives/2008/03/`, 2008.

[63] D. Seltzer-Kelly, S. J. Westwood, and D. M. Peña-Guzman. A methodological self-study of quantitizing: Negotiating meaning and revealing multiplicity. *Journal of Mixed Methods Research*, 6(4):258–274, 2012.

[64] S. Shuai, D. Guowei, G. Tao, Y. Tianchang, and

S. Chenjie. Modelling analysis and auto-detection of cryptographic misuse in Android applications. In *Proceedings of the 12th IEEE International Conference on Dependable, Autonomic, and Secure Computing*, pages 75–80, Aug 2014.

[65] Stack Exchange. Stack overflow. `https://stackoverflow.com/`, 2018.

[66] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *USENIX Security Symposium*, pages 399–416, 2009.

[67] A. S. Tanenbaum. *Computer Networks: 2Nd Edition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[68] The MITRE Corporation. Common vulnerabilities and exposures (CVE). `https://cve.mitre.org/`, 2018.

[69] D. R. Thomas. A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2):237–246, June 2006.

[70] Underwriters Laboratories (UL). Certification. `https://services.ul.com/categories/certification/`, 2018.

[71] US-CERT. OpenSSL Heartbleed vulnerability (CVE-2014-0160). `https://www.us-cert.gov/ncas/alerts/TA14-098A`, 2014.

[72] Veracode. The state of software security. `https://www.veracode.com/sites/default/files/Resources/Reports/state-of-software-security-volume-7-veracode-report.pdf`, 2016.

[73] Veracode. Veracode secure development survey: Developers respond to application security trends. `https://info.veracode.com/report-veracode-developer-survey.html`, 2016.

[74] L. Williams, R. R. Kessler, W. Cunningham, and R. Jeffries. Strengthening the case for pair programming. *IEEE Software*, 17(4):19–25, 2000.

[75] S. Xiao and E. Witschey, J.and Murphy-Hill. Social influences on secure development tool adoption: Why security tools spread. In *Proceedings of the 17th ACM conference on Computer supported Cooperative Work and Social Computing*, CSCW '14, pages 1095–1106, Feb. 2014.

[76] J. Xie and B. Lipford, H. R.and Chu. Why do programmers make security errors? In *Proceedings of the 2011 IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 161–164, Sept. 2011.

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

APPENDIX

A. INTERVIEW QUESTIONS

1. Can you tell me about your organization - what it does, what it produces?
2. What is your role within your organization with respect to cryptographic products?
3. How did you get into this field?
   (a) At what point and why did you become concerned with cryptography and secure development?
   (b) In which field(s) is your formal education?
4. Do you work in a unit or department that is part of a larger organization?
   If yes : What is the size of the unit or department?
   (a) What is the size of your overall organization?
5. Can you tell me about the kinds of products your organization develops, and specifically those that use cryptography?
6. Who are the typical customers for your products that use cryptography?
7. How long has your organization been working on products that use cryptography?
8. Is cryptography your organization's primary business focus, or is it an enabler within your products?
9. For your products that use cryptography, what processes or techniques , if any, does your organization use to minimize bugs and errors in code during the development process?
   (a) Why does your organization choose to use these methods? [only use if participant has difficulty coming up with response:] for example, industry standard, customer demand, robustness and quality
10. What processes or techniques does your organization use to test and validate the cryptography component in your products?
    (a) Why does your organization choose to use these methods? [only use if participant has difficulty coming up with response:] for example, industry standard, customer demand, robustness and quality
    (b) What kind of end-user testing, if any, does your organization do to prevent customers from misconfiguring or misusing the cryptography component in your products?
11. Does your organization do any certifications or third-party testing?
    (a) What reasons led you to decide to use certifications or third-party testing?
    (b) How do you establish confidence in the results of the certifications or third-party testing?
    (c) What are the challenges or issues your organization has experienced with certifications or third-party testing, if any?
12. What, if any, are your organization's biggest challenges with respect to developing and testing cryptography within your products?
    (a) How do you think these challenges can be overcome, if at all?
    (b) Has your organization experienced a tension between secure development and testing and getting a product to market? If so, how has that impacted your organization's processes?
13. Do your customers have specific requirements regarding development and testing? If so, what are those requirements?
14. How do updates impact your development and testing processes, if at all? (time-sensitive vs. deprecation)
15. What resources do you use to help you develop and test the cryptography component of your products? [only use if participant has difficulty coming up with response:] for example, standards, industry specifications, books, academic papers, standard libraries, APIs
    (a) What are the reasons your organization chooses to use those particular resources?
    If the participant does NOT use standards: What are the reasons that your organization does not use standards?
16. [If the participant uses standards:] What kinds of standards do you use?
    (a) What is the role of standards in your organization's development and testing processes?
    (b) What do you see as the value or benefit of using these standards, if any?
17. How could standards or other cryptographic resources be improved to be more useful?
    (a) How could NIST standards and guidance be improved to be more useful?
18. Is there anything else you'd like to add about the topics we've discussed?

Haney, Julie; Theofanos, Mary; Acar, Yasemin; Prettyman, Sandra.
""We make it a big deal in the company": Security Mindsets in Organizations that Develop Cryptographic Products."
Paper presented at Fourteenth Symposium on Usable Privacy and Security, Baltimore, MD, United States. August 12, 2018 - August 14, 2018.

B. CODEBOOK

Participant Demographics

Organization Demographics

Organization Characteristics

- Team Interactions
- Security Culture
- Maturity
- Talent/Hiring

Development and Testing Practices

- Formal
- Informal
- Risk-based
- Practices
  - Automated
  - Human/Manual
  - External/3rd party testing
  - End-user testing
- Reasons/Philosophy
- Evolution
- Confidence
- Challenges
- Updates

Certifications/Compliance Programs

- Which ones (identify)
- Problems and Challenges
- Reasons
- Confidence
- Improvements

Resources

- Standards
- Government
- Industry/3rd Party
- Internally developed
- Research
- Gaps

Security

- Vulnerabilities and Errors
- Usability and Complexity
- Relationship and Tensions

Security Education

- Customers
- Developers/Engineers

Emotions

- Positive
- Negative

Influences

Customers

Evolution of Security Field

Complaints

Participant Values and Perceptions

Trust

A Portable Cold ^{87}Rb Atomic Clock with Frequency Instability at One Day in the 10^{-15} Range

F. G. Ascarrunz, Y. O. Dudin, Maria C. Delgado Aramburo and L. I. Ascarrunz
SpectraDynamics, Inc.
Louisville, USA

J.Savory, Alessandro Banducci and S. R. Jefferts
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO USA

***Abstract*—We present a small portable cold ^{87}Rubidium atomic clock with frequency uncertainty of less than $\mathbf{3 \times 10^{-15}}$ in one day of averaging time. The clock is under development at SpectraDynamics for the Defense Advanced Research Projects Agency (DARPA) Portable Microwave Cold Atomic Clock Small Business Innovation Research (SBIR) effort. The portable clock is about the size of a desktop computer (22x37x32 cm) and weighs 28kg.**

***Keywords*—laser-cooled atomic clock, atomic clock, Rb clock**

I. Introduction

We present a small, portable, laser-cooled ^{87}Rb atomic clock with frequency uncertainty of less than $\mathbf{3 \times 10^{-15}}$ in one day of averaging time. The clock is under development at SpectraDynamics for the Defense Advanced Research Projects Agency (DARPA) Portable Microwave Cold Atomic Clock Small Business Innovation Research (SBIR) effort. The portable clock is about the size of a desktop computer (22x37x32 cm) and weighs 28kg. If we compare to a commercial Cs beam clock with the high-performance option, we note that the volume of our clock is nearly identical, with the Cs clock being 29.74l and our Rb clock having a volume of 26l. The mass of both clocks is also very similar with the Cs clock being 30kg compared to the Rb clock at 28kg. The one-day instability of the Rb clock is about a factor of 10 smaller than a high-stability commercial Cs beam clock.

This clock is intended to be a product and is targeted for release into both defense and commercial markets. The prototype clock has outputs at 1 pulse-per-second, 5 MHz, 10 MHz and 100 MHz.

II. The Clock

The clock uses ^{87}Rb atoms which are trapped and cooled using a compact 780nm laser system. Following state preparation, the light is extinguished and the atoms enter a microwave cavity where they undergo microwave interrogation. After the microwave interrogation sequence, the fraction of atoms that made the clock transition is counted. This signal is then used to control the frequency output of the clock. The complete clock-cycle is shown in *Figure 2* and the clock block diagram is depicted in *Figure 3*.

The clock, which has been operational for about one year, typically operates continuously unless deliberately interfered

Figure 1 – *Prototype of the portable cold Rb87 atomic clock. The clock is packaged in a 22 X 37 X 32 cm enclosure and weighs about 28 kg. The signal and power connections are on the rear panel of the instrument.*

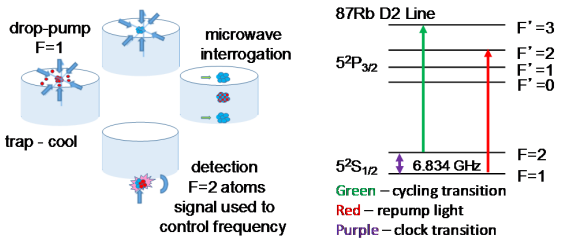


Figure 2 – *The clock cycle and the 87Rb D2 line transitions used for the clock.*

This research was, in part, funded by the U.S. Government. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Savory, Joshua; Ascarrunz, F; Ascarrunz, L; Banducci, Alessandro; Delgado Aramburo, M; Dudin, Y; Jefferts, Steven.
"A Portable Cold 87Rb Atomic Clock with Frequency Instability at One Day in the $10^{u} - 15^{R}ange.$"
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

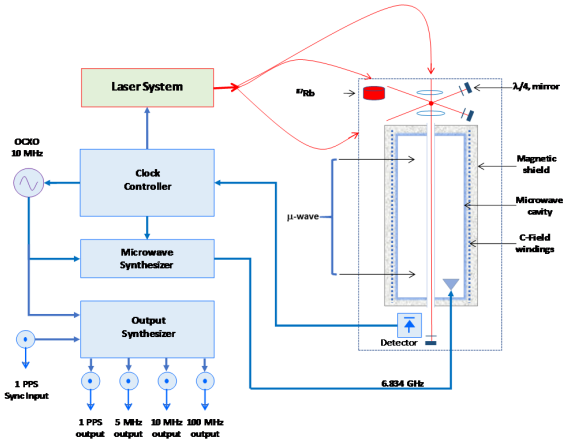


Figure 3 – *Block diagram of the portable cold Rb87 atomic clock illustrating all major subsystems.*

with by an operator. The clock has operated in excess of 100 days with no interference. The 100 day-long run was stopped by the operators to investigate other aspects of clock operation.

III. MEASUREMENT SETUP

The prototype clock has been compared to the time-scale UTC(NIST) for extended periods, allowing characterization of both the short-term stability $\left(\sigma_y(\tau) < 8 \times 10^{-13}/\sqrt{\tau}\right)$, and the long-term behavior of the clock [5]. The comparison is made between the clock's 5 MHz output and UTC(NIST) using the same measurement system which collects the clock data for the generation of UTC(NIST) [1,2,3]. During the same period of time, a second 5 MHz output of the clock is compared to a high-stability Hydrogen maser [4]. The measurement was configured to sample at 1 Hz allowing a characterization of the very short-term performance of the clock.

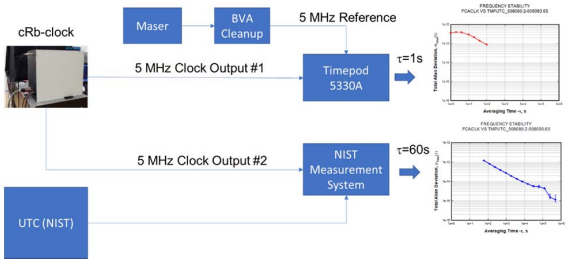


Figure 4 – *Clock measurement system configuration.*

IV. MEASUREMENT RESULTS

As shown below in *Figure 5*, the short-term performance of the clock, from $\tau = 1\ s$ out to $\tau \cong 10{,}000\ s$, is characterized by an Allan Deviation [5] of $\sigma_y(\tau) \cong 8 \times \frac{10^{-13}}{\sqrt{\tau}}$. The short-term performance can be improved further at the cost of a more

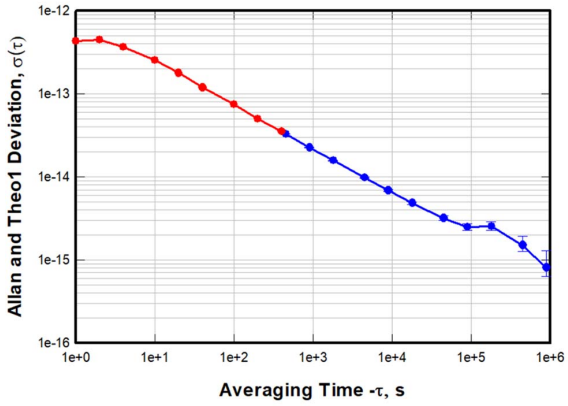


Figure 5 – *Total Allan Deviation of the cold Rb clock as measured against UTC(NIST) [5]. No drift has been removed. The 1 s data is shown in Red, while the 60 s data is shown in Blue. The last two points in Blue come from the Theo1 statistic.*

expensive local oscillator, but, even at this level, the clock should be able to reach a frequency uncertainty of less than 3×10^{-15} in one day of averaging. The long-term frequency instability is more complicated, in general, the Total Allan Deviation [6] plot shows a characteristic flattening of the previous $1/\sqrt{\tau}$ slope starting at around a day and persisting out to slightly longer than two days. The clock begins to improve with further averaging at that point. This behavior is illustrated in *Figure 5*.

Another way of comparing the cold-Rb clock is to look at other clocks in the UTC (NIST) timescale over the same measurement period. In *Figure 6* we show the frequency fluctuations of several clocks as measured by the NIST measurement system. Plotted in Figure 6, in Green is a very low drift maser, while a more typical maser is shown in Red. The cold-Rb clock is shown in Dark Blue, while a commercial high-performance Cs beam clock is shown in Light Blue (Cyan).

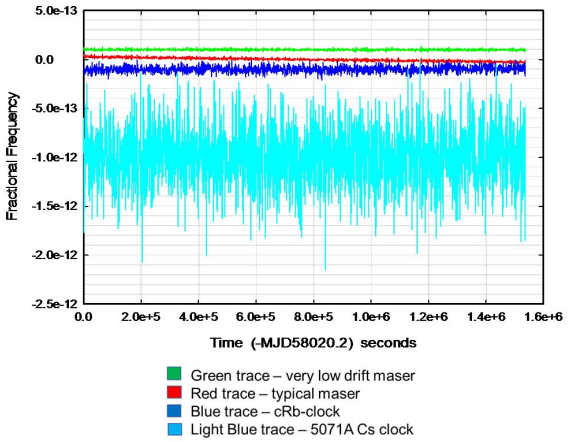


Figure 6 – *Several clocks measured simultaneously by the NIST measurement system [5].*

Savory, Joshua; Ascarrunz, F; Ascarrunz, L; Banducci, Alessandro; Delgado Aramburo, M; Dudin, Y; Jefferts, Steven.
"A Portable Cold 87Rb Atomic Clock with Frequency Instability at One Day in the $10^{u} - 15^{R}ange.$"
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

Several things are readily apparent: first, the frequency fluctuations of the cold-Rb clock are much smaller than those of the commercial Cs clock albeit, not as small as the frequency fluctuations in a good hydrogen maser. Second, comparing the Blue curve to the Red curve, it is apparent that the drift in the cold-Rb clock is much smaller than the drift in a typical maser. In fact, the residual drift in the cold Rb clock is about 1×10^{-17}/day, consistent with no drift. All of the clocks shown in *Figure 6* have had arbitrary frequency offsets added to separate the curves for illustration purposes.

In *Figure 7* we show the Allan deviation of several active Hydrogen masers as measured by the NIST measurement system. The Allan deviation of the cold-Rb clock, shown in Blue, indicates better clock performance than all of the five masers for average times greater than 6 days.

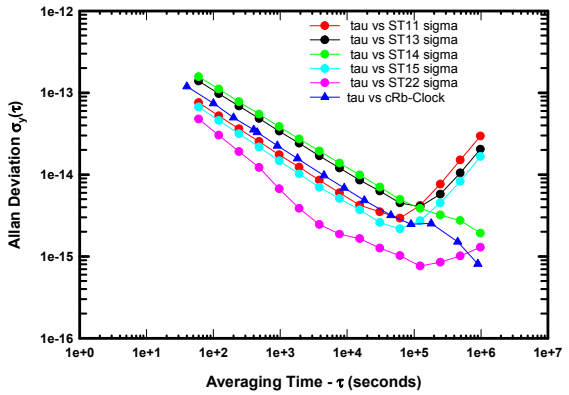


Figure 7 – *Several active Hydrogen masers measured by the NIST measurement system. In the long term the cold-Rb clock outperforms all of these masers.*

V. CONCLUSIONS

We present a new class of portable atomic clock with short-term frequency instability from 1 to 10,000 s, characterized by $\sigma_y(\tau) \cong 8 \times \frac{10^{-13}}{\sqrt{\tau}}$, and a long-term frequency instability of less than 9×10^{-16}. The frequency drift of the clock is less than 1×10^{-17}/day.

ACKNOWLEDGMENT

The authors thank Drs. Robert Lutwak and John Burke for their support of the program. This work was funded in part by the DARPA under contract HR0011-15-9-0007.

REFERENCES

[1] Judah Levine. Invited review article: The statistical modeling of atomic clocks and the design of time scales. Review of Scientific Instruments, 83(2):021101, 2012.

[2] S. Römisch, S. R Jefferts, and T. E. Parker, "A digital time scale at the National Institute for Standards and Technology," in General Assembly and Scientific Symposium, 2011 XXXth URSI (IEEE, 2011), pp. 1–4.

[3] S. Stein, D. Glaze, J. Gray, D. Hilliard, D. Howe, and L. A. Erb, 1983, "Automated high accuracy phase measurement system, " IEEE Transactions on Instrumentation and Measurement 32, 227-231.

[4] H. E. Peters and H. B. Owings "Hydrogen maser improvements and future applications," in Proc. IEEE Int. Freq. Contr. Symp., 997, pp. 280-285.

[5] Ascarrunz, F. G., Dudin, Y. O., Aramburo, Maria. C. Delgado, Savory, J., Jefferts, S.R., "Long-Term Frequency Instability of a Portable Cold 87Rb Atomic Clock," Proceedings of the 49th Annual Precise Time and Time Interval Systems and Applications Meeting, Reston, Virginia, January 2018, pp. 107-110.

Savory, Joshua; Ascarrunz, F; Ascarrunz, L; Banducci, Alessandro; Delgado Aramburo, M; Dudin, Y; Jefferts, Steven.
"A Portable Cold 87Rb Atomic Clock with Frequency Instability at One Day in the $10^{u} - 15^{Range}$."
Paper presented at 2018 IEEE International Frequency Control Symposium (IFCS), Olympic Valley, CA, United States. May 21, 2018 - May 24, 2018.

Nanoscale Imaging of Photocurrent in Perovskite Solar Cells using Near-field Scanning Photocurrent Microscopy

Dongheon Ha[1, 2], Yohan Yoon[1, 2], Ik Jae Park[3], Paul M. Haney[1], and Nikolai B. Zhitenev[1]

[1]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland, 20899, USA

[2]Maryland Nanocenter, University of Maryland, College Park, Maryland, 20742, USA

[3]Department of Materials Science and Engineering, Seoul National University, Seoul, 151-742, Korea

ABSTRACT

We study photocurrent generation and collection of methylammonium lead iodide perovskite solar cells with nanoscale resolution using a near-field scanning photocurrent microscopy (NSPM) technique. For NSPM measurements, we employ a non-contact mode atomic force microscopy probe with an attached optical fiber coated with Cr/Au metal. We observe an increased photocurrent at grain boundaries in samples annealed at moderate temperature (100 °C); however, the opposite spatial pattern is observed in samples annealed at higher temperature (130 °C). Combining the NSPM results with other characterization techniques such as electron microscopy, X-ray diffraction, and quantum efficiency measurements, we show that the cause of the photocurrent contrast is the material inhomogeneity and the dynamics of lead iodide. The NSPM technique is further used to establish the mechanism of the cell degradation under extended light illumination.

I. INTRODUCTION

Due to their easy fabrication, low-cost, and surprisingly fast improvement in the power conversion efficiency, perovskite solar cells have attracted significant interest in the photovoltaic (PV) research community [1], [2]. Recent developments offer a pathway for high efficiency solar cells without costly tandem structures [3], [4] and/or nanophotonic engineering [5]–[11]. However, various instability issues caused by environmental factors (*e.g.*, extended light illumination, humidity, temperature, etc.) have impeded their deployment as commercial solar cells. The as of yet unknown factors causing degradation should be firmly established.

Multiple measurement techniques have been applied for the characterization of material composition and structural properties of perovskite solar cells [12]–[15] at the micro- or nanoscale. However, the detailed studies connecting the degradation with the nanoscale photo-excited carrier generation and collection are still lacking.

In this study, we image nanoscale photocurrent collection of perovskite solar cells by a near-field scanning photocurrent microscopy (NSPM) technique. Correlating the nanoscale photocurrent images with other characterization techniques (X-ray diffraction, electron microscopy, quantum efficiency measurements, etc.), we relate the cell preparation temperature with its operation at the macro-/nanoscale. We also determine how the light-induced degradation affects the cell microstructure and the local carrier collection by monitoring the changes in the NSPM photocurrent images which occur during the experiment (on the timescale of hours).

II. RESULTS AND DISCUSSION

For the NSPM measurements, we employ a tuning fork-based non-contact mode atomic force microscopy (AFM) probe with an attached multimode optical fiber to locally inject light [9], [11], [16]. The end of the optical probe is coated with Cr (20 nm) and Au (200 nm) metal, and the probe has an output hole with a diameter of ≈ 200 nm (see

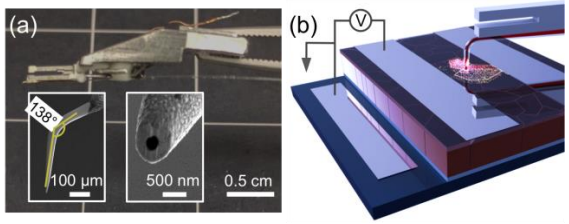


Fig. 1. (a) An SEM image showing an AFM probe used for NSPM measurements. Metal-coated optical fiber used for near-field light injection is attached to a tuning fork. Insets: enlarged side-view of the probe (left), enlarged view of the probe at the bottom (right). (b) A schematic (not to scale) describing NSPM measurements on perovskite solar cells. Photocurrent amplified by a variable gain amplifier is detected with a lock-in technique.

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoscale imaging of photocurrent in perovskite solar cells using near-field scanning photocurrent microscopy."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

Fig. 1a). The perovskite solar cell is placed onto a piezo stage of an AFM system, and topography and photocurrent signals are simultaneously obtained (see Fig. 1b). The NSPM probe has a force constant of 15 N/m and a resonance frequency of ≈ 35 kHz. A diode laser with a wavelength of 635 nm is connected to the optical fiber of the NSPM probe via a FC/PC connector. A variable-gain low-noise current amplifier is used for signal amplification, and the amplified photocurrent is detected with a lock-in amplifier. During the raster scanning, the distance between the NSPM probe and the top surface of the perovskite solar cell is kept ≈ 10 nm. In some measurements, we used different output hole sizes of the NSPM probe (between 50 nm and 300 nm) or changed the laser power to generate a measurable photocurrent or to avoid the degradation during the measurements.

Our perovskite solar cell has methylammonium lead iodide ($CH_3NH_3PbI_3$, $MAPbI_3$ hereafter) as its active layer. It is enclosed by [6,6]-phenyl-C61-butyric acid methyl ester (PCBM) and nickel oxide (NiO_x) as electron and hole transport layers, respectively. The full device structure is shown in Fig. 2a, and the device consists of silver (Ag) contact/PCBM/400 nm thick perovskite layer ($MAPbI_3$)/NiO_x/indium tin oxide (ITO), from top to bottom.

First, we determine macroscopic electrical properties of perovskite solar cells annealed at two different temperatures: 100 °C and 130 °C (see Fig. 2b). The sample annealed at moderate temperature (100 °C) exhibits superior photovoltaic properties (*i.e.*, short circuit current density, J_{SC}, open circuit voltage, V_{OC}, and fill factor, *FF*) compared to the sample annealed at higher temperature (130 °C). The power conversion efficiency is η = 16.98 % for the sample annealed at 100 °C vs. η = 14.81 % for the sample annealed at 130 °C. Repeated measurements under identical conditions yield maximum variation of less than 1 % [11], [16].

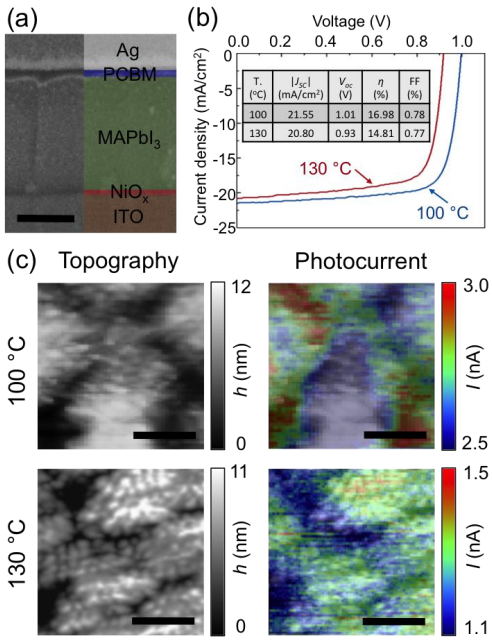


Fig. 2. (a) Device structure of perovskite solar cells. Scale bar is 200 nm. (b) Current-Voltage characteristics of perovskite solar cells annealed at two different temperatures. The sample annealed at moderate temperature (100 °C, blue line) exhibits higher power conversion efficiency than the sample annealed at higher temperature (130 °C, red line). (c) Nanoscale photocurrent imaging of perovskite solar cells using the NSPM technique. Opposite photocurrent collection patterns are observed between the two samples. Scale bars are 500 nm.

The difference between the cells is clearly seen in NSPM photocurrent images (see Fig. 2c). The nanoscale photocurrent patterns are exactly opposite for the two samples annealed at different temperatures: the photocurrent is higher at grain boundaries for the sample annealed at 100 °C while it is higher at grain interiors for the sample annealed at 130 °C. The beneficial role of grain boundaries in carrier generation and collection is attributed to the dynamics of lead iodide (PbI_2) segregation [16], [17]. In the sample annealed at 100 °C, a moderate amount of PbI_2 segregates at grain boundaries leading to the defect passivation. However, more structural and compositional transformation from $MAPbI_3$ to PbI_2 takes place with further annealing, resulting in distinguishable crystalline phases of PbI_2 at grain boundaries in the sample annealed at 130 °C. This leads to an increased grain boundary recombination, and grain interiors become more efficient carrier collectors. Note that the measured photocurrent is not caused by surface morphology or thickness variation of the perovskite solar cells, as confirmed by the significant contrast in the measured photocurrents (up to ≈ 20 % and ≈ 36 % in the sample annealed at 100 °C and 130 °C, respectively).

As mentioned above, the instability and the performance degradation under operational environment is the major issue for perovskite solar cell technology. We study how perovskite solar cells change their operation under extended light illumination. First, we identify the microscopic material and compositional changes of perovskite solar cells using X-ray diffraction measurement and electron microscopy (Fig. 3a). Measurements are performed when samples are aged under the Air Mass 1.5 Global (AM1.5G) spectrum with an intensity of 10 mW/cm^2 (0.1 sun illumination). The moderate light dose is selected to avoid any possible degradation due to overheating. A relative humidity is kept ≈ 25 % at room temperature during the whole measurements.

The diffraction peak corresponding to PbI_2 is observed at a diffraction angle of 12.75° (marked with #) for both samples annealed at different temperatures. The intensity of this peak increases as the cell ages, indicating continuous material decomposition from $MAPbI_3$ to PbI_2. PbI_2 crystallites are clearly distinguishable in cross-

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoscale imaging of photocurrent in perovskite solar cells using near-field scanning photocurrent microscopy."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

sectional scanning electron microscopy (SEM) image of the sample annealed at 130 °C. Distinct spots with bright contrasts are observed at grain boundaries due to their low conductivity and increased accumulation of charges (see yellow dotted lines in Fig. 3a). However, as seen in SEM images, the distribution of PbI_2 is quite different in two samples annealed at different temperatures: PbI_2 is evenly distributed in the active area for the sample annealed at moderate temperature (100 °C), and it is concentrated at grain boundaries for the sample annealed at higher temperature (130 °C). However, as the aging time increases, the distribution of PbI_2 in the sample annealed at 100 °C is no longer uniform over the sample, showing distinguishable phases of PbI_2 in SEM images.

The external quantum efficiency (EQE) of the samples is determined as a function of illumination time (see Fig. 3b). Minor changes are observed for 200 min of light degradation, however, both samples show significant deterioration as aging time increases. The sample annealed at higher temperature (130 °C) collects more photo-generated carriers than the sample annealed at 100 °C for the same aging time. Repeated measurements under identical conditions yield maximum variation of less than 1 % [11], [16], [18].

We investigate how the extended light illumination affects nanoscale carrier generation and collection by monitoring the changes in the NSPM photocurrent images during the aging process (see Fig. 4). As shown in Fig. 2c, photocurrent patterns are opposite for two samples annealed at two different temperatures: grain boundaries are more efficient carrier collectors for the sample annealed at moderate temperature (100 °C), while they are less efficient than grain interiors for the sample annealed at the higher temperature (130 °C). Even though the photo-generated current is reduced under the aging process, this carrier collection pattern persists for the sample annealed at higher temperature (130 °C). However, the sample annealed at moderate temperature (100 °C) exhibits quite different behavior. At 400 min of aging process, the photocurrent collection at the grain boundary marked with a white dotted-circle is strongly suppressed. The photocurrent was significantly higher at this spot in pristine condition (t = 0). The loss of photocurrent enhancement with aging for this sample is attributed to the dynamics of PbI_2 segregation.

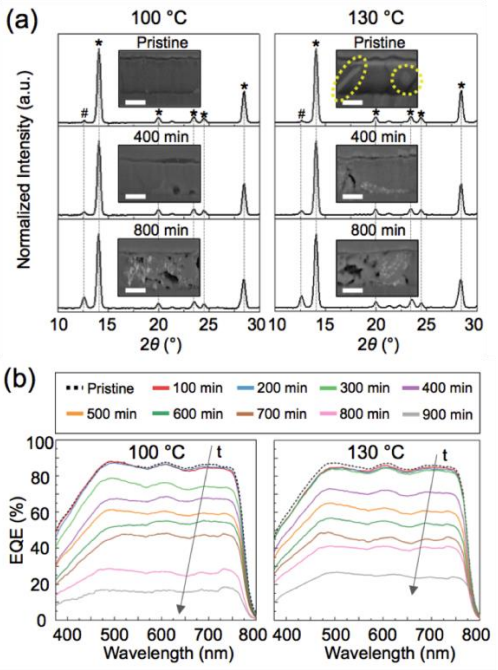


Fig. 3. (a) X-ray diffraction patterns for perovskite solar cells under light-induced aging process. # and * denote the identified diffraction peaks corresponding to PbI_2 and $MAPbI_3$, respectively. Yellow dotted lines show distinguishable PbI_2 crystallites at grain boundaries. Inset SEM images show cross-sectional view of samples under the light-induced aging process. Scale bars are 200 nm. (b) EQEs of perovskite solar cells under the light-induced aging process.

As seen in Fig. 3a, more structural decomposition from $MAPbI_3$ to PbI_2 occurs under the continuous light illumination, and this leads to a growth of distinguishable PbI_2 crystallites at grain boundaries, somewhat similar to the pristine state of the sample annealed at higher temperature. The enhanced carrier collection at grain boundaries is no longer observed for this sample. The suppression of the enhanced collection at grain boundaries significantly affects the overall device performance. This can explain the faster degradation of this sample during the aging process (see Fig. 3b). For the other sample, the main paths for carrier generation and collection remain the same, resulting in relatively robust photovoltaic properties under aging process, as seen from the macroscopic EQE measurements in Fig. 3b.

IV. Conclusions

We study nanoscale photocurrent of methylammonium lead iodide perovskite solar cells using a near-field scanning photocurrent microscopy. We show that the spatial pattern of photocurrent is related to the sample preparation temperature: higher photocurrent is observed at grain boundaries in samples annealed at moderate temperature (100 °C), while lower

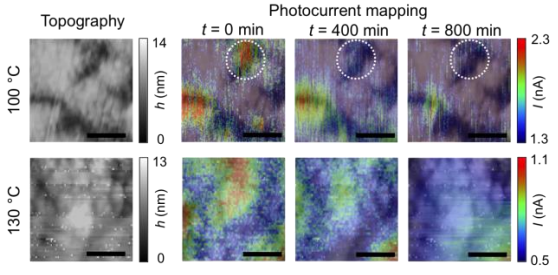


Fig. 4. NSPM-based nanoscale photocurrent images of perovskite solar cells at multiple stages of light-induced aging. Scale bars are 500 nm.

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoscale imaging of photocurrent in perovskite solar cells using near-field scanning photocurrent microscopy."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

photocurrent is observed at grain boundaries and higher photocurrent is observed at grain interiors in samples annealed at higher temperature (130 °C). Correlating nanoscale photocurrent images with other characterization techniques, we show that the particular spatial patterns of photocurrent are due to the material inhomogeneity and the dynamics of segregation of lead iodide. We also determine how the light-induced aging process affects the nanoscale carrier collection in perovskite solar cells. The structural and compositional changes of materials through the aging process suppress the beneficial role of grain boundaries in samples annealed at moderate temperature.

ACKNOWLEDGEMENTS

The authors thank S. Lee for fruitful discussions of data analysis. D. Ha and Y. Yoon acknowledge support under the Cooperative Research Agreement between the University of Maryland and the Center for Nanoscale Science and Technology at the National Institute of Standards and Technology, Award 70NANB14H209, through the University of Maryland.

REFERENCES

[1] M. A. Green, Y. Hishikawa, W. Warta, E. D. Dunlop, D. H. Levi, J. Hohl-Ebinger, and A. W. H. Ho-Baillie, "Solar cell efficiency tables (version 50)," *Prog. Photovolt. Res. Appl.,* vol. 25, pp. 668-676, 2017.
[2] M. A. Green, A. Ho-Baillie, and H. J. Snaith, "The emergence of perovskite solar cells," *Nat. Photonics*, vol. 8, pp. 506-514, 2014.
[3] R. R. Kinga, D. C. Law, K. M. Edmondson, C. M. Fetzer, G. S. Kinsey, H. Yoon, R. A. Sherif, and N. H. Karam, "40% efficient metamorphic GaInP⁄GaInAs⁄Ge multijunction solar cells," *Appl. Phys. Lett.*, vol. 90, 183516, 2007.
[4] L. Dou, J. You, J. Yang, C. -C. Chen, Y. He, S. Murase, T. Moriarty, K. Emery, G. Li, and Y. Yang, "Tandem polymer solar cells featuring a spectrally matched low-bandgap polymer," *Nat. Photonics*, vol. 6, pp. 180-185, 2012.
[5] D. Ha, Z. Fang, L. Hu, and J. N. Munday, "Paper-based anti-reflection coatings for photovoltaics," *Adv. Energy Mater.*, vol. 4, 1301804, 2014.
[6] Z. Fang, H. Zhu, Y. Yuan, D. Ha, S. Zhu, C. Preston, Q. Chen, Y. Li, X. Han, S. Lee, G. Chen, T. Li, J. Munday, J. Huang, L. Hu, "Novel nanostructured paper with ultrahigh transparency and ultrahigh haze for solar cells," *Nano Lett.*, vol. 14, pp. 765-773, 2014.
[7] D. Ha, J. Murray, Z. Fang, L. Hu, and J. N. Munday, "Advanced broadband antireflection coatings based on cellulose microfiber paper," *IEEE J. Photovolt.*, vol. 5, pp. 577-583, 2015.
[8] D. Ha, C. Gong, M. S. Leite, and J. N. Munday, "Demonstration of resonance coupling in scalable dielectric microresonator coatings for photovoltaics," *ACS Appl. Mater. Interfaces*, vol. 8, pp. 24536-24542, 2016.
[9] D. Ha, Y. Yoon, and N. B. Zhitenev, "Nanoscale imaging of photocurrent enhancement by resonator array photovoltaic coatings," *Nanotechnology*, vol. 29, 145401, 2018.
[10] D. Ha, Z. Fang, and N. B. Zhitenev, "Paper in electronic and optoelectronic devices," *Adv. Electron. Mater.*, vol. 4, 1700593, 2018.
[11] D. Ha and N. B. Zhitenev, "Improving dielectric nano-resonator array coatings for solar cells," *Part. Part. Syst. Charact.* vol. 35, 1800131, 2018.
[12] J. S. Yun, A. Ho-Baillie, S. Huang, S. H. Woo, Y. Heo, J. Seidel, F. Huang, Y. -B. Cheng, and M. A. Green, "Benefit of grain boundaries in organic–inorganic halide planar perovskite solar cells," *J. Phys. Chem. Lett.*, vol. 6, pp. 875-880, 2015.
[13] J. H. Kim, P. -W. Liang, S. T. Williams, N. Cho, C. -C. Chueh, M. S. Glaz, D. S. Ginger, and A. K. -Y. Jen, "High-performance and environmentally stable planar heterojunction perovskite solar cells based on a solution-processed copper-doped nickel oxide hole-transporting layer," *Adv. Mater.*, vol. 27, pp. 695-701, 2015.
[14] V. W. Bergmann, S. A. L. Weber, F. J. Ramos, M. K. Nazeeruddin, M. Grätzel, D. Li, A. L. Domanski, I. Lieberwirth, S. Ahmad, and R. Berger, "Real-space observation of unbalanced charge distribution inside a perovskite-sensitized solar cell," *Nat. Commun.*, vol. 5, 5001, 2014.
[15] J. L. Garrett, E. M. Tennyson, M. Hu, J. Huang, J. N. Munday, and M. S. Leite, "Real-time nanoscale open-circuit voltage dynamics of perovskite solar cells," *Nano Lett.*, vol. 17, pp. 2554-2560, 2017.
[16] Y. Yoon, D. Ha, I. J. Park, P. M. Haney, S. Lee, and N. B. Zhitenev, "Nanoscale photocurrent mapping in perovskite solar cells," *Nano Energy*, vol. 48, pp. 543-550, 2018.
[17] Q. Chen, H. Zhou, T. -B. Song, S. Luo, Z. Hong, H. -S. Duan, L. Dou, Y. Liu, and Y. Yang, "Controllable Self-Induced Passivation of Hybrid Lead Iodide Perovskites toward High Performance Solar Cells," *Nano Lett.*, vol. 14, pp. 4158-4163, 2014.
[18] D. Ha, "Microscale dielectric anti-reflection coatings for photovoltaics," Ph.D. Thesis, University of Maryland, 2016.

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoscale imaging of photocurrent in perovskite solar cells using near-field scanning photocurrent microscopy."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

Contributed Papers from Materials Science and Technology 2018 (MS&T18)
October 14–18, 2018, Greater Columbus Convention Center, Columbus, Ohio, USA

DOI 10.7449/2018/MST_2018_1334_1341

NEAR-SURFACE ELEMENTAL ANALYSIS OF SOLIDS BY NEUTRON DEPTH PROFILING

Jamie L. Weaver and R. Gregory Downing

Materials Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

Keywords: Neutrons, Depth Profiling, solids analysis, materials characterization

Abstract

Neutron Depth Profiling (NDP) is an analytical nuclear technique for the determination of elemental mass as a function of depth into a sample's surface. It is suitable for characterizing a few select elements in most solid materials to depths of micrometers and with the resolution of 10's of nanometers. The most commonly measured elements are helium, lithium, boron and nitrogen. Applications include the study of implanted boron in semiconductor substrates, measurement of helium embedded into first shielding wall alloys/metals used in fusion reactors, lithium diffusion through all solid-state thin-film batteries, and nitrogen penetration into corroded building materials. A brief review of past NDP research as well as NDP instrumentation development both at NIST and other national and international NDP research facilities will be presented. Future research opportunities and current development projects will be discussed.

Introduction

Neutron depth profiling (NDP) is a non-destructive, analytical nuclear technique for the determination of elemental mass as a function of its depth in a sample's surface. The method on which NDP is based was first described in 1972 by Zeigler *et al.* to study B impurities in solids [1]. It was advanced to its current state by Biersack and coworkers (Institute Laue-Langevin, Grenoble) in the 70's and 80's [2, 3]. Currently, there are more than seven (7) NDP instruments in operation around the globe [4], which are maintained by neutron science user facilities.

Methods

The key reaction of NDP is neutron capture by a nucleus to form an unstable, compound nucleus, which will undergo an exoergic charged particle reaction (Figure 1). The reaction products are either an alpha particle or a proton, and a recoiling nucleus. The emitted recoil particles have distinct initial energies - which are indicators of their parent nucleus. As the energetic particles move through the solid they will predominately lose energy as the result of interactions with electrons in the material. The rate of energy loss per unit distance from the particle is a function of the elemental composition and volumetric density of the material, and it is determined by the material's characteristic stopping power. The difference in the initial energy of the charged particles and their final energies as they exit the solid is directly related to the depth of the reacted nucleus from which the particles originated. The relationship between origin depth and measured energy of the particle can be expressed as:

$$x = \int_{E_f}^{E_i} \frac{dE}{S(E)} \quad (1)$$

Where x is the path length traveled by the particle through the solid, E_i is the initial energy of the particle, E_f is the energy of the particle upon exiting the solid, and $S(E)$ is the stopping power of the material.

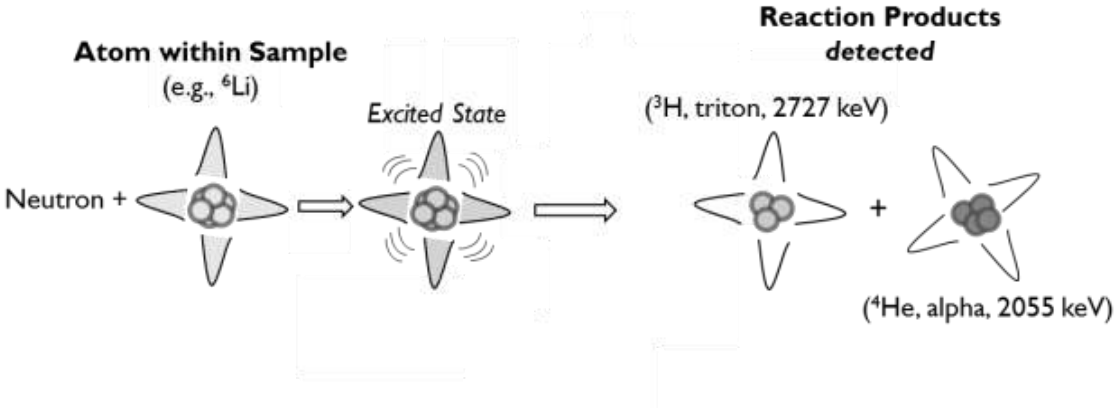


Figure 1. Nuclear reaction for NDP. Depicted example is the reaction of a neutron (n) with ^{6}Li.

Applications

There are only a few select isotopes for which this reaction can be used to measure depth profiles in the near-surface of solids; the most common of which are ^{3}He, ^{6}Li, ^{10}B, and ^{14}N. Other isotopes which have been studied include ^{17}O, ^{22}Na, and ^{7}Be. Much of the early research using NDP was on ^{10}B (^{10}B(n,α)^{7}Li) in semiconductors and metals [5]. ^{10}B is the isotope of boron measured by NDP, which has a very large thermal neutron cross-section (3400 and 241 barns for its two branching reactions), and is 19.9% naturally abundant [6]. NDP has also been utilized to measure B distribution in glasses [7, 8], as well as boron bearing carbides [9].

Li, like B, is another popular element measured by NDP. This is due, in part, to the relative insensitivity of other techniques (particularly those which are x-ray based) to Li. ^{6}Li is the isotope measured by NDP, with a relatively large thermal neutron cross-section (941 barns), and is 7.59% naturally abundant [6]. In recent years the measurement of Li distribution in Li-ion battery materials by NDP has dominated NDP research. *Ex-situ (for examples see* [10, 11]), *in-situ* (for example see [12]), and *in-operando* (for example see [13]) studies of Li diffusion through battery materials and whole cells have been completed by NDP. NDP results coupled with other materials characterization and electrochemical techniques has enabled determination of Li transport numbers through solid-state cells [14], Li diffusion coefficients [15], effect of manufacturing parameters on cell stability and performance (see Figure 2), and to determine the feasibility of novel battery materials [16, 17]. Multi-dimensional mapping of Li in battery electrodes has been achieved within the last decade [18], and there is much anticipation of its continued development. Additionally, advances in lateral spatial mapping of Li in solids have been made using multipixel detectors originally designed for use at CERN [19]. Materials in

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

which Li profiles have been measured by NDP include ceramics [20], optical wave guides, polymers, and glasses [21].

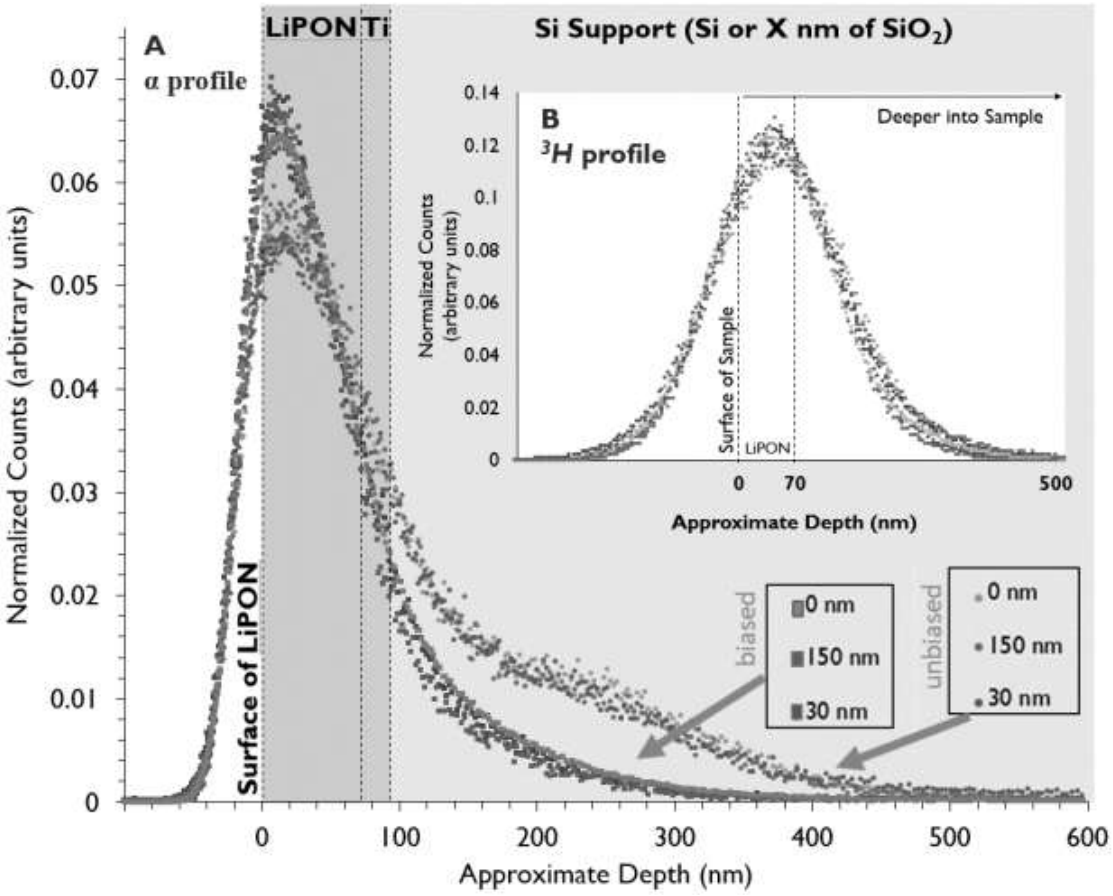


Figure 2. ^{6}Li neutron depth profile collected on a series of solid-state battery cells manufactured under different sputtering conditions (electrically biased and unbiased). NDP allowed for measurement of Li out-diffusion from the cell constructed with a glassy electrolyte (LiPON) as a function of manufacturing and post-processing parameters. Data collected on NIST-NDP instrument.

^{3}He research by NDP has focused on the measurements of implantation profiles of the isotope in alloys and metals. In the late 80's ^{3}He implantation into nickel and amorphous nickel materials was used to understand the microscopic structure and diffusion characteristics through solids[22]. Recently, ^{3}He profiling experiments have been completed for the investigation of shield walls planned for use in next generation fusion reactors. A significant challenge facing the long-term operation of test fusion reactors is the durability of plasma-facing materials and components to intense neutron/neutral/ion particle bombardment and implantation [23]. Tritium is one such erosive particle, and it is known to cause surface blistering and spallation due to entrapment in surface of metals and alloys. ^{3}He is the daughter product of tritium beta decay (½

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

life of 12.32 years). When tritium is implanted by plasma into the materials its penetration depth and blistering/exposure rates can be measured once it decays to the NDP-detectable ^{3}He ($^{3}He(n,p)^{3}H$) [24]. Similar irradiation damage by helium has been recently measured in $MgAl_2O_4$ crystals, which is a possible matrix to be used for the transmutation and storage of americium [25].

Implantation of ^{14}N into metals has also been investigated by NDP. Over the last few decades there has been considerable interest in the nitride coatings and implantation of nitrogen into metals as means of forming corrosion resistant surface layers, and into semi-conductors and polymers for electronic applications. Fink and co-workers determined through their NDP studies of ^{14}N doped into metals and alloys at high fluence rates that the isotope would diffuse through Mg and Al at room temperature directly following implantation, and through other metals at greater time lengths (0.5 to 4 years) [26]. Of the materials studied, only Cu, Si, and W retained a small percentage of the implanted N over time, while all other materials (Be, Al, Ni, stainless steel, Ta, and Cd) lost all implanted N by diffusion. Other areas of research where ^{14}N NDP may be applicable include sensors and biotechnology [27], as well as archaeometry, and materials corrosion.

Instrument Variations

Sample preparation and NDP measurements are relatively straight forward for the investigation of isotopic profiles in solids. Analysis time is statistically dependent upon the neutron fluence rate and the analyte concentration within the depth interval of charged particle escape from the surface. That is, profile depth resolution is largely a function of analysis time, neutron fluence rate, and mass fraction of the analyte.

The sample should have a relatively smooth surface over the area to be analyzed. The area is not required to have a regular perimeter as an aperture can be customized to fit the shape and area of the surface. The mass fraction of the analyte is determined by analyzing a well-characterized standard using the same aperture and taking a ratio of the known mass to the unknown sample. Calculations are normalized for run time and for total neutron fluence using a neutron monitor simultaneously with the sample analysis. Time dependent measurements are possible by obtaining a series of spectra at appropriate intervals.

All measurements are obtained in an evacuated chamber. The vacuum is needed to ensure that charged particles emerging from the sample surface do not lose additional energy as they travel between the sample surface and the charged particle detector. All NDP instruments are custom built and no two systems are designed the same. Typically, they are designed to best function at the neutron beam line being used. The NDP system in use at NIST is shown in Figure 3. The beam exits the neutron guide at the far right. Neutrons pass a short distance in air and enters the NDP instrument through a thin aluminum window. The beam passes through the chamber containing the sample and exits the instrument through a thin aluminum window. Finally, the residual neutrons continue into a beam absorber at the far left as an aluminum encased box containing lithium and boron rich materials to stop the neutrons and lead shielding to block any gamma component.

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

Samples are introduced through the lid at the top of the chamber when the system is at atmospheric pressure. There are numerous ports that radiate from the sides and bottom of the stainless-steel chamber. These ports serve to attach vacuum lines, vacuum sensors, up to air values, and electrical connections for the detectors, motion controllers to the sample and detectors, and any electrical connections required for sample environment controls. These environmental controls could dynamically charge and discharge battery systems, heat or cool the sample, attach to thermocouples, translate the sample through the beam, or any variety of research control needs. [28-31]

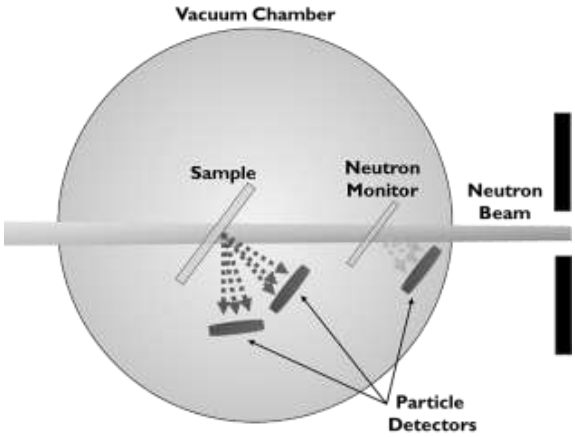


Figure 3. Diagram depicting the inside of the NDP vacuum chamber. All NDP instruments nominally contain the same functionality. The neutron beam traverses the vacuum chamber intersecting the sample mounted near the center of the system. A sample will typically set at an angle to the incoming beam and directly face a charged particle detector placed several centimeters from the sample surface and away from the neutron beam. In the NIST NDP system a neutron beam monitor sits at entrance to the chamber; however other facilities can have the monitor outside the front of the chamber.

Samples are mounted to a removeable holder that can be reproducibly located in the chamber for comparison to other samples, controls, blanks, and standards. A custom-cut aperture is fixed to the sample holder and all of the sample is masked except the area to be analyzed. Neutrons pass through both the masked area and the aperture but the induced charged particles can only pass through the aperture opening to the detector. The charged particles reaching the detector, or detectors, do so with a small angular spread out of the sample such that particles from equal sample depth travel the same distance through the sample.

A variety of charged particle detectors have been used in NDP systems. The silicon surface-barrier detector (SSB) exhibits the best combination of simplicity, energy resolution, and minimal background from gamma-ray, x-ray, and other non-signal sources. Other systems include PIN, ion implanted, time-of-flight, and pixelated detectors. PIN detectors are becoming the most widely used detector. They are inexpensive relative to the SSB detectors and have

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

comparable energy resolution but are susceptible to considerable spectral noise that starts around 1400 keV and increases logarithmically at lower energies. Time-of-flight (ToF) detectors have the potential for the highest energy resolution and lowest background in NDP measurement [32]. However, ToF detector systems used in NDP environment are notoriously difficult to implement and maintain thus they are not in common use. The pixelated detector systems [19] have much potential for 2-, 3-, and 4-dimensional measurement, where the 4th dimension is time. Count rate and energy resolution is not comparable to the other systems but are quite useful for multidimensional applications.

Finally, NDP detection can be performed in coincidence mode [33]. NDP reactions produce two energetic masses of characteristic initial energy and those two particles always are emitted diametrically from the reaction site. For samples that are thin enough that both particles can escape the sample and the sample has parallel faces, the energy-depth resolution is exceptional. This is because of mathematical constraints that can be placed upon the coincident signals [34]. Also, there is no restriction upon the acceptance angle at the detector, so both detectors can be placed as close to the sample as practical and remain out of the neutron beam. The limitation on the sample dimensions and cost limit the applicability of coincidence NDP.

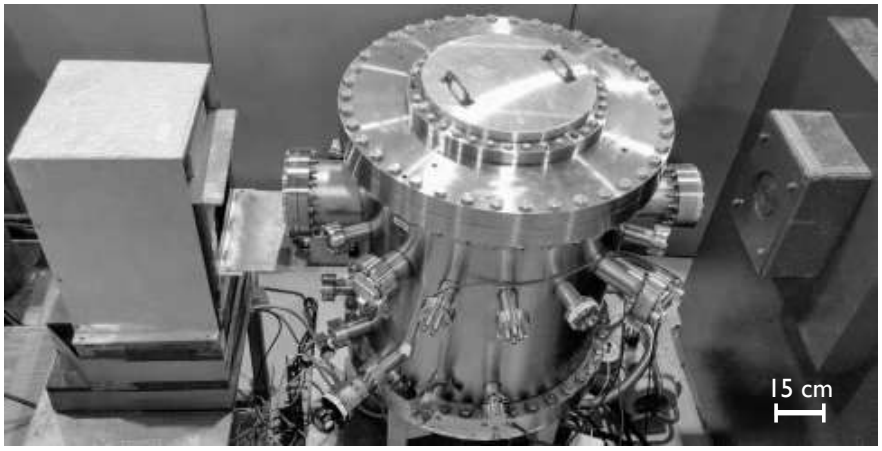


Figure 4. *NDP chamber setup at NG-5 at NIST*. *Left* is neutron beam stop, *center* is the NDP chamber, and *right* is exit of neutron guide.

Conclusion

Neutron depth profiling is a non-destructive, near-surface analyses method which is sensitive to several light elements. Its applications have varied since its conception nearly 50 years ago, with application to the fields of nuclear physics, solid-state chemistry, semiconductor materials, corrosion science, tribology, battery science, and many more. Details of various NDP systems and the compliment of NDP applications can be found through the papers given in reference section.

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

Acknowledgements and Disclaimer

The authors would like to thank the NIST Center for Neutron Research (NCNR) and Center for Nanoscale Science and Technology (CNST) at NIST for their technical support. Partial funding for this research was provided by the United States National Research Council. Contributions of the National Institute of Standards and Technology are not subject to copyright.

References

1. Ziegler, J., G. Cole, and J. Baglin, *Technique for determining concentration profiles of boron impurities in substrates.* Journal of Applied Physics, 1972. **43**(9): p. 3809-3815.
2. Biersack, J., et al., *Range profiles and thermal release of helium implanted into various metals.* Journal of Nuclear Materials, 1979. **85**: p. 1165-1171.
3. Biersack, J., et al., *The use of neutron induced reactions for light element profiling and lattice localization.* Nuclear instruments and Methods, 1978. **149**(1-3): p. 93-97.
4. Downing, R.G. *Historical List of Neutron Depth Profiling Facilities*. 2018 [cited 2018 06/01/2018]; Available from: https://sites.google.com/site/nistndp/earlierfacilitieslist.
5. Biersack, J. and D. Fink, *Implantation of Boron and Lithium in Semiconductors and Metals*, in *Ion Implantation in Semiconductors*. 1975, Springer. p. 211-218.
6. Baum, E.M., et al., *Nuclides and Isotopes: Chart of the Nuclides*. 17 ed. 2010: Knolls Atomic Power Laboratory. 94.
7. Chen-Mayer, H.H., G.P. Lamaze, and S.K. Satija. *Characterization of BPSG films using Neutron Depth Profiling and Neutron/X-ray Reflectometry*. in *AIP Conference Proceedings*. 2001. AIP.
8. Riley, J., et al. *Materials Analysis Using Thermal Neutron Reactions: Applications*. in *Materials Science Forum*. 1984. Trans Tech Publ.
9. Clayton, D.A., J.L. Lacy, and R.G. Downing, *Quality assurance in neutron detector development by neutron depth profiling.* Transactions of the American Nuclear Society, 2006. **95**: p. 397-398.
10. Lamaze, G.P., et al., *Analysis of lithium transport in electrochromic multilayer films by neutron depth profiling.* Surface and Interface Analysis: An International Journal devoted to the development and application of techniques for the analysis of surfaces, interfaces and thin films, 2000. **29**(9): p. 637-641.
11. Lamaze, G.P., et al., *Cold neutron depth profiling of lithium-ion battery materials.* Journal of power sources, 2003. **119**: p. 680-685.
12. Oudenhoven, J., et al., *In Situ Neutron Depth Profiling: A Powerful Method to Probe Lithium Transport in Micro-Batteries.* Advanced Materials, 2011. **23**(35): p. 4103-4106.
13. Lv, S., et al., *Operando monitoring the lithium spatial distribution of lithium metal anodes.* Nature Communications, 2018. **9**(1): p. 2152.
14. Liu, D.X., et al., *In situ quantification and visualization of lithium transport with neutrons.* Angewandte Chemie International Edition, 2014. **53**(36): p. 9498-9502.
15. Nagpure, S.C., et al., *Neutron depth profiling technique for studying aging in Li-ion batteries.* Electrochimica Acta, 2011. **56**(13): p. 4735-4743.
16. Liu, D.X., L.R. Cao, and A.C. Co, *Demonstrating the Feasibility of Al as Anode Current Collector in Li-Ion Batteries via In Situ Neutron Depth Profiling.* Chemistry of Materials, 2016. **28**(2): p. 556-563.
17. Han, X., et al., *Negating interfacial impedance in garnet-based solid-state Li metal batteries.* Nature materials, 2017. **16**(5): p. 572.
18. He, Y., R.G. Downing, and H. Wang, *3D mapping of lithium in battery electrodes using neutron activation.* Journal of Power Sources, 2015. **287**: p. 226-230.
19. Tomandl, I., et al., *High resolution imaging of 2D distribution of lithium in thin samples measured with multipixel detectors in sandwich geometry.* Review of Scientific Instruments, 2017. **88**(2): p. 023706.

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

20. McWhinney, H.G., et al., *Diffusion of lithium-6 isotopes in lithium aluminate ceramics using neutron depth profiling.* Journal of nuclear materials, 1993. **203**(1): p. 43-49.
21. Trivelpiece, C.L., et al., *Glass Surface Layer Density by Neutron Depth Profiling.* International Journal of Applied Glass Science, 2012. **3**(2): p. 137-143.
22. Uenleu, K., *Helium-3 in nickel-base amorphous metals: Surface features, subsurface microstructure, migration, and release upon annealing*. 1989, Michigan Univ., Ann Arbor, MI (USA).
23. Shimada, M. and B. Merrill, *Tritium decay helium-3 effects in tungsten.* Nuclear Materials and Energy, 2017. **12**: p. 699-702.
24. Gilliam, S.B., et al., *Retention and surface blistering of helium irradiated tungsten as a first wall material.* Journal of Nuclear Materials, 2005. **347**(3): p. 289-297.
25. Neeft, E.A.C., et al., *Helium irradiation effects in single crystals of MgAl2O4.* Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 2000. **166-167**: p. 238-243.
26. Fink, D., et al., *Nitrogen depth profiling using the 14N (n, p) 4C reaction.* Nuclear Instruments and Methods in Physics Research, 1983. **218**(1-3): p. 171-175.
27. Fink, D., et al., *Nuclear track-based biosensing: an overview.* Radiation Effects and Defects in Solids, 2016. **171**(1-2): p. 173-185.
28. Downing, R., et al., *Neutron depth profiling: overview and description of NIST facilities.* Journal of research of the National Institute of Standards and Technology, 1993. **98**(1): p. 109.
29. Ünlü, K. and B.W. Wehring, *Neutron depth profiling at the University of Texas.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1994. **353**(1-3): p. 402-405.
30. Vacik, J., et al., *Neutron depth profiling facility at Nuclear Physics Institute Rez.* Acta Physica Hungarica, 1994. **75**(1): p. 369-372.
31. Downing, R.G., *NIST Neutron Depth Profiling Facility: 2013, invited.* Transactions-American Nuclear Society, 2013. **109**.
32. Çetiner, S., K. Ünlü, and R. Downing, *Development of time-of-flight neutron depth profiling at Penn State University.* Journal of radioanalytical and nuclear chemistry, 2007. **271**(2): p. 275-281.
33. Parikh, N.R., et al., *Neutron depth profiling by coincidence spectrometry.* Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 1990. **45**(1-4): p. 70-74.
34. Trivelpiece, C.L. and J. Brenizer, *Development of a geometric uncertainty model describing the accuracy of position-sensitive, coincidence neutron detection.* Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 2011. **269**(1): p. 45-52.

Weaver, Jamie; Downing, Robert.
"Near-Surface Elemental Analysis of Solids by Neutron Depth Profiling."
Paper presented at MS&T Conference Proceedings, Columbus, OH, United States. October 14, 2018 - October 18, 2018.

Sesame: A Numerical Simulation Tool for Polycrystalline Photovoltaics

Benoit Gaury[1,2], Yubo Sun[3], Peter Bermel[3], and Paul Haney[1]

Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA)

Maryland NanoCenter, University of Maryland, College Park, MD 20742, USA

School of Electrical & Computer Engineering, Purdue University, West Lafayette, IN, USA

***Abstract* — We present a new software simulation tool "Sesame", which solves the drift-diffusion-Poisson equations in 1 and 2-dimensions. Sesame is distributed both as an open source Python package and as a standalone executable for Windows. The software is designed to enable easy construction of systems with extended defects such as grain boundaries and sample surfaces. In this paper, we present an overview of Sesame's capabilities along with results benchmarking Sesame against commercial software packages. The source code, executable, and full documentation with tutorials is available at https://pages.nist.gov/sesame**

***Index Terms* — numerical simulation, polycrystalline solar cell, software.**

I. Introduction

Numerical simulations play a central role in photovoltaic development and research. Freely available software packages for describing solar cells in 1-dimension have been widely used for decades. Among these include AMPS [1], PC-1D [2], SCAPS [3], wxAMPS [4]. A 1-dimensional description is adequate for systems which are translationally invariant along the directions perpendicular to the *p-n* junction interface, such as single crystal Si or GaAs. However, there are a number of photovoltaic materials with lateral inhomogeneities, such as thin film polycrystalline materials including CdTe and CIGS. The grain boundaries in these materials result in complex geometries that require 2 or 3-dimensional modeling, and which play a key role in device performance [5]-[6]. Free simulation software for 2-dimensional systems is relatively less commonly available, which motivates the development and release of a new software package "Sesame". Sesame is an open source Python package recently developed by the authors (B. G and P. M. H.), and includes documentation and tutorials. In this paper we describe some details of Sesame's usage and benchmark its results relative to commercial software packages. The source code, executable, and full documentation with tutorials is available at https://pages.nist.gov/sesame.

II. Program Description

Sesame solves the drift-diffusion-Poisson equation in 1 and 2-dimensions, and includes Shockley-Read-Hall, radiative, and Auger recombination mechanisms. Sesame supports variable electronic structure and can describe heterojunctions or graded materials. Sesame is currently limited to describing non-degenerate semiconductors with Boltzmann statistics, and does not include thermionic emission and quantum tunneling at interfaces. These can be important contributions to the transport in heterojunctions [7], so care should be exercised when using Sesame to simulate such systems. Sesame includes Ohmic and Schottky contact boundary conditions, and periodic or hardwall (infinite potential) transverse boundary conditions. Sesame uses finite differences to solve the continuity and Poisson equations, and the standard Scharfetter-Gummel scheme for discretizing the current [8]. Sesame is designed to easily construct systems with planar defects, such as grain boundaries or sample surfaces, which may contain both discrete or a continuum of gap state defects. This enables a description of grain boundaries in polycrystalline solar cells, and the ability to model experiments for which sample surface effects are important (such as Electron Beam Induced Current or Scanning Kelvin Probe Microscopy). Sesame is open source and is distributed under the BSD license.

A. Python scripting

As a Python package, Sesame can be called within any Python script. This mode of use enables users to generate complex systems (e.g. materials with graded electronic properties) and to perform large scale "batch" calculations, where many simulations are distributed on a computing cluster. The number of simulations increases exponentially with number of varied parameters, but access to a cluster enables the efficient computation of many system configurations. Access to a large number of solutions in parameter space is particularly useful for determining the important nonlinearities which control the system behavior. In the distribution, we provide several example scripts which describe standard PV simulations (e.g. J-V, IQE calculations), along with in-depth tutorials. Sesame also includes tools for efficiently saving and loading the results of simulations, and tools for data analysis and plotting.

B. Graphical User Interface

Gaury, Benoit; Sun, Yubo; Bermel, Peter; Haney, Paul.
"Sesame: A Numerical Simulation Tool for Polycrystalline Photovoltaics."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

To facilitate Sesame's use by those without Python distributions, we've compiled Sesame into a standalone executable for Windows with a graphical user interface using the PyQt library. Use of the standalone GUI as an alternative to scripting can be more convenient for small-scale calculations. Simulation settings can be saved and loaded, and the standalone GUI also includes a full Python distribution which can be accessed with an interactive Python console.

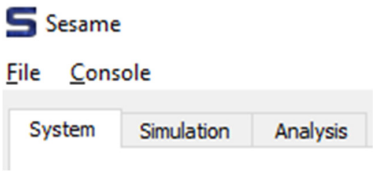


Fig. 1. Menu and table structure of the Sesame GUI.

The GUI is divided into three tabs: 1. The "System" tab contains fields to define the system geometry and material parameters. Fig. 2 shows a portion of the system tab where planar defects are defined. 2. The "Simulation" tab lets the user specify which parameter is varied: either the voltage is swept, or a user-defined variable related to the generation rate density is swept. The boundary conditions and output file information are also set here, and the simulation is launched from this tab. A window on this tab provides details on the calculation progress. 3. The "Analysis" tab enables the user to plot the output of the simulation, and to save and export plotted data. Sesame is distributed with several sample input files for setting up standard PV simulations in the GUI. More detailed documentation for the GUI is included in the distribution.

The standalone executable contains a full Python distribution, and the scripting capability can be accessed by selecting the "Console" option on the main menu (see Fig. 1). This opens an interactive Python console, from which Python scripts can be executed. In addition, the numerical output generated by scripts can be loaded into the GUI for plotting and analysis. The capability of the GUI to run scripts and analyze/plot the output of scripts allows for a flexible usage of the Sesame package, which can be optimized according to the user's needs and experience.

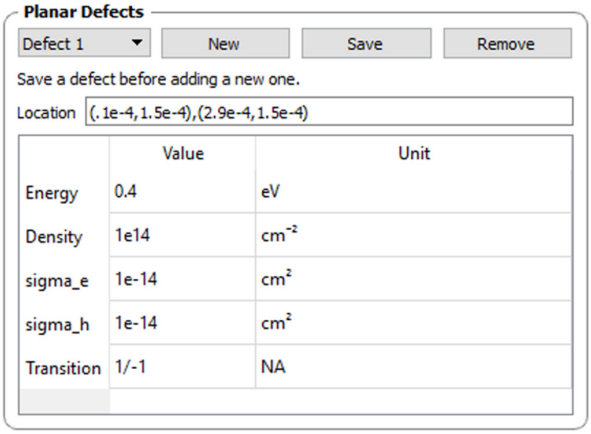


Fig. 2. Portion of the "System" tab of Sesame GUI for specifying planar defects.

III. BENCHMARKS

To verify the accuracy of Sesame, we compared its results with that of Sentaurus [9] and COMSOL's Semiconductor module [10]-[11]. We first consider a 2-dimensional *p-n* homojunction with a single columnar grain boundary (see inset of Fig. 3(a) for a depiction of the geometry). The bulk material parameters are given in Table 1. The *n*-type region is taken to have a thickness of 100 nm with doping 10^{17} cm^{-3}, while the p-type region has thickness 2.9 μm with doping 10^{15} cm^{-3}. The grain boundary parameters are given in Table 1. The grain boundary contains both a donor and an acceptor defect at the given energy level. The contacts are taken to be Ohmic for both majority and minority carries, and we use hardwall transverse boundary conditions. The generation rate density is given as $G(x) = \phi_0 \exp(-\alpha x)$, where $\phi_0 = 10^{17}$ (cm$^2 \cdot$ s)$^{-1}$, $\alpha = 2.3 \times 10^4$ cm^{-1}.

Fig. 3(a) shows the comparison of the illuminated J-V curve computed with Sesame and Sentaurus. To quantify the comparison, we define the relative difference between two computed currents J_1 and J_2 as $|J_1 - J_2|/((J_1 + J_2)/2)$. We find excellent agreement between software packages, with a relative difference of less than 1 % in the computed current. Fig. 3(b) shows the computed band structure along the grain boundary core under short circuit conditions, where we find that the two solutions coincide.

Gaury, Benoit; Sun, Yubo; Bermel, Peter; Haney, Paul.
"Sesame: A Numerical Simulation Tool for Polycrystalline Photovoltaics."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

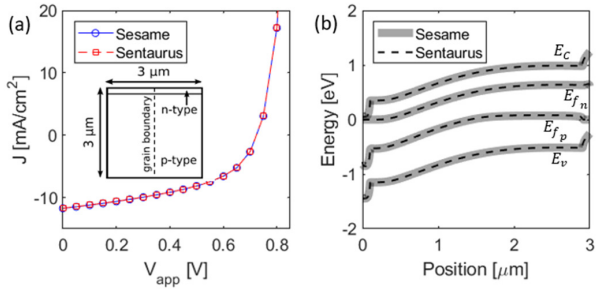


Fig. 3. Comparison between Sesame and Sentaurus for a 2-dimensional homojunction system. (a) is the J-V curve obtained with given software packages under nonuniform (see inset for geometry). (b) is the band diagram along the grain boundary core at short-circuit conditions obtained by the given software packages.

We next consider a 2-d system with an additional grain boundary, which is oriented at an oblique angle with respect to the *p-n* junction, and which intersects the initial columnar grain boundary (see inset of Fig. 4(a) for geometry, and the "GB2" row of Table 1 for exact position of the additional grain boundary). We use this system to compare the output of Sesame and COMSOL Semiconductor Module. For this system, we find a less than 2 % difference between Sesame and COMSOL Semiconductor Module in the computed current values. Fig. 4(b) shows the J-V curve on a log scale.

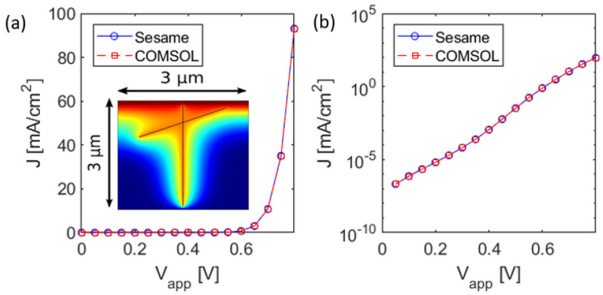


Fig. 4. Comparison between Sesame and COMSOL Semicoductor Model for a 2-dimensional homojunction system, with two grain boundaries. The inset of (a) shows a colormap representation of the electrostatic potential of the system in equilibrium, together with lines indicated the grain boundary positions. (a) is the J-V curve obtained with given software packages under dark conditions. (b) shows the J-V curve on a log scale.

The two examples given here are intended to indicate the veracity and functionality of Sesame. We emphasize that there are several important questions that can be addressed with this tool, such as determining the impact of grain boundary networks on the device efficiency of polycrystalline PV. Previous work has considered the role of grain boundaries on the open-circuit voltage [13], but has not addressed the short-circuit current density or fill factor. It is likely that understanding the impact of grain boundaries on the overall efficiency will require substantial additional analysis.

TABLE I
DEFAULT SIMULATION PARAMETERS

Parameter [units]	Value
ϵ	9.4
τ_n [ns]	10
τ_p [ns]	10
N_C [cm^{-3}]	8×10^{17}
N_V [cm^{-3}]	1.8×10^{19}
E_g [eV]	1.5
χ [eV]	3.9
μ_n [cm^2/(V · s)]	320
μ_p [cm^2/(V · s)]	40
Grain Boundary Parameters	
Defect density [cm^{-2}]	10^{14}
Defect energy [eV]	0.4 (above midgap)
σ_e [cm^2]	10^{-14}
σ_h [cm^2]	10^{-14}
GB1 endpoints (x,y) [μm]	(1.5, 0.1), (1.5, 2.9)
GB2 endpoints (x,y) [μm]	(2.5, 0.2), (0.5, 1.0)

IV. CONCLUSION

We present a free, open source distribution of photovoltaic device modeling software. It is available freely as a standalone executable and as a Python package. We believe that the study of complex geometries of defects within *p-n* junctions contains more physics to uncover, and hope that Sesame is a tool that opens this field to a wide class of researchers. The source code is open and documented, and can be easily modified and supplemented according the terms of the BSD license.

REFERENCES

[1] S. Fonash, et al., "A Manual for AMPS-1D for Windows 95/NT", The Pennsylvania State University, 1997.
[2] P. A. Basore and D. A. Clugston, "PC1D Version 4 for Windows: From Analysis to Design", *25th IEEE Photovoltaic Specialists Conference*, Washington, May 1996, pp.377- 381.
[3] M. Burgelman, P. Nollet and S. Degrave, "Modelling polycrystalline semiconductor solar cells", *Thin Solid Films*, vol. 361-362, pp. 527-532, 2000.
[4] Liu, Yiming, Yun Sun, and Angus Rockett, "A new simulation software of solar cells—wxAMPS." *Solar Energy Materials and Solar Cells*, vol. 98, pp. 124-128, 2012.
[5] S. Edmiston, G. Heiser, A. Sproul, and M. Green, "Improved modeling of grain boundary recombination in bulk and p - n junction regions of polycrystalline silicon solar cells." *Journal of Applied Physics*, vol. 80, pp. 6783-6795, 1996.
[6] M. Gloeckler, J. R. Sites, and W. K. Metzger, " Grain-boundary recombination in Cu(In,Ga)Se2 solar cells." *Journal of Applied Physics,* vol. 98, pp. 113704 1-10, 2005.
[7] K. Horio and H. Yanai, " Numerical modeling of heterojunctions including the thermionic emission mechanism at the

Gaury, Benoit; Sun, Yubo; Bermel, Peter; Haney, Paul.
"Sesame: A Numerical Simulation Tool for Polycrystalline Photovoltaics."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

heterojunction interface." *IEEE Transactions on Electron Devices*, vol. 37, pp. 1093-1098, 1990.
[8] H. K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculations." *IEEE Transactions on Electron Devices*, vol. 11, pp. 455-465, 1964.
[9] S. D. U. Guide and E. Version, Mountain View, CA, USA, 2013.
[10] C. Multiphysics, COMSOL AB, Stockholm, Sweden, 2017.
[11] The full description of the procedures used in this paper requires the identification of certain commercial products. The inclusion of such information should in no way be construed as indicating that such products are endorsed by NIST or are recommended by NIST or that they are necessarily the best software for the purposes described.
[12] J. D. Major, "Grain boundaries in CdTe thin film solar cells: a review." *Semiconductor Science and Technology,* vol. 31, p. 093001, 2016.
[13] B. Gaury and P. M. Haney, "*Charged grain boundaries reduce the open-circuit voltage of polycrystalline solar cells—An analytical description*", Journal of Applied Physics vol. 120, p. 234503, 2016.

Gaury, Benoit; Sun, Yubo; Bermel, Peter; Haney, Paul.
"Sesame: A Numerical Simulation Tool for Polycrystalline Photovoltaics."
Paper presented at 45th IEEE Photovoltaic Specialists Conference (7th World Conference on Photovoltaic Energy Conversion), Waikoloa, HI, United States. June 10, 2018 - June 15, 2018.

Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort

Roselyne B. Tchoua[1], Aswathy Ajith[1], Zhi Hong[1], Logan T. Ward[2], Kyle Chard[2,3], Alexander Belikov[6], Debra J. Audus[4], Shrayesh Patel[5], Juan J. de Pablo[5], and Ian T. Foster[1,2,3]

[1] Department of Computer Science, University of Chicago, Chicago, IL, USA
roselyne@uchicago.edu
[2] Globus, University of Chicago, Chicago, IL, USA
[3] Data Science & Learning Division, Argonne National Lab, Lemont, IL, USA
[4] Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
[5] Institute for Molecular Engineering, University of Chicago, Chicago, IL, USA
[6] Knowledge Lab, University of Chicago, Chicago, IL, USA

Abstract. Scientific Named Entity Referent Extraction is often more complicated than traditional Named Entity Recognition (NER). For example, in polymer science, chemical structure may be encoded in a variety of nonstandard naming conventions, and authors may refer to polymers with conventional names, commonly used names, labels (in lieu of longer names), synonyms, and acronyms. As a result, accurate scientific NER methods are often based on task-specific rules, which are difficult to develop and maintain, and are not easily generalized to other tasks and fields. Machine learning models require substantial expert-annotated data for training. Here we propose polyNER: a semi-automated system for efficient identification of scientific entities in text. PolyNER applies word embedding models to generate entity-rich corpora for productive expert labeling, and then uses the resulting labeled data to bootstrap a context-based word vector classifier. Evaluation on materials science publications shows that the polyNER approach enables improved precision or recall relative to a state-of-the-art chemical entity extraction system at a dramatically lower cost: it required just two hours of expert time, rather than extensive and expensive rule engineering, to achieve that result. This result highlights the potential for human-computer partnership for constructing domain-specific scientific NER systems.

Keywords: Scientific Named Entities · Word Embedding · Natural Language Processing · Crowdsourcing · Polymers.

1 Introduction

There is a pressing need for automated information extraction and machine learning (ML) tools to extract knowledge from the scientific literature. One task that such tools must perform is the identification of entities within text. Many

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

rule-based, ML, and hybrid named entity recognition (NER) approaches have been developed for particular entity types (e.g., people and places) [23, 20].

Scientific NER remains challenging due to non-standard encoding and the use of multiple entity *referents* (terms used to refer to an entity). For example, in materials science, polymers are encoded in text using various representations (conventional or commonly-used), acronyms, synonyms, and historical terms. Further challenges arise when trying to distinguish between general and specific references to members of polymer families, or recognizing references to blends of two polymers, etc. Such challenges are not unique to polymer science. However, while NER is a well-studied topic in medicine and biology [5, 16], it has only recently become a focus in materials science [10, 29, 17, 36]. Many approaches applied in other domains rely on large, carefully annotated corpora of training data, a luxury not yet available in domains like polymer science.

Here we introduce polyNER, a hybrid computer-human system for semi-automatically identifying scientific entity referents in text. PolyNER operates in three phases, first applying a fully automated analysis to produce an entity-rich set of candidates for labeling; then engaging experts to approve or reject a modest number of proposed candidates; and finally using the resulting labeled candidates to train a classifier. In both the first and third phases, it uses word embedding models to capture shared contexts in which referents occur. PolyNER thus seeks to substitute the labor-intensive processes of either assembling a large manually labeled corpus or defining complex domain-specific rules with a mix of sophisticated automated analysis and focused expert input.

We evaluate polyNER performance on polymer science publications. We compare the output of its first candidate enrichment phase against expert-labeled data, and find that it retrieves 61.2% of the polymers extracted by experts with a precision (26.0%) far higher than the ratio of target entities vs. non-entities in scientific publications (less than 2% in our experience). We evaluate the performance of polyNER overall by comparing its output polymer referents against both expert-labeled data and a state-of-the art rule-based chemical entity extraction system, ChemDataExtractor (CDE) [36], which we have previously enhanced with dictionary- and rule-based methods for identifying polymers [39]. We find that PolyNER can achieve either 52.7% precision or 90.7% recall, depending on user preference, a 10.5% improvement in precision or 22.4% improvement in recall over the enhanced CDE. These results highlight the potential for creating domain-specific scientific NER systems by combining sophisticated automated analysis with focused expert input.

The rest of this paper is as follows. We review scientific NER systems in Section 2. In Section 3, we motivate the need for identifying polymer names in text. We describe design and implementation in Section 4 and evaluate polyNER in Section 5. We summarize and discuss future work in Section 6.

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

2 Related Work

Natural Language Processing (NLP) is a way for computers to "read" human language. NLP tasks include automatic summarization, topic modeling, translation, named entity recognition (NER), and relationship extraction. NER systems, which aim to identify and categorize named entities (e.g., a person, organization, location), have been developed using both linguistic grammar-based techniques and ML models. While many ML models have been developed, their performance depends critically on the quantity and quality of training data.

Scientific domains such as molecular biology and medical NLP have long used NER for extracting symptoms, diagnoses, medications, etc. from text [5, 16]. More recently, there has also been much interest in chemical entity and drug recognition [10, 29, 17, 36]. However, even state-of-the-art NER systems do not typically perform well when applied to different domains [13]. Considerable effort is involved in selecting and (often manually) generating quality data for trainable statistical NER systems [14].

Various approaches have been proposed to address the lack of training data for NER and other information extraction tasks. *Distant supervision* maps known entities and relations from a structured knowledge base onto unstructured text [26, 43]. However, many fields, including polymer science, lack such knowledge bases.

Data programming uses *labeling functions* (user-defined programs that provide labels for subsets of data) [27]. Errors due to differences in accuracy and conflicts between labeling functions are addressed by learning and modeling the accuracies of the labeling functions. Under certain conditions, data programming achieves results on par with those of supervised learning methods. But while writing concise scripts to define rules may seem to be a more reasonable task for annotators than exhaustively annotating text, it still requires expert guidance. Moreover, labeling functions typically rely on state-of-the-art entity taggers, such as CoreNLP [19], which recognizes persons, locations, organizations and more, and which itself has been trained using various corpora, including the Conference on Computational Natural Language Learning (CoNLL) dataset [32]. A user-defined function may be defined, for example, as: *if the word "married" appears between two PERSONs (as identified by a state-of-the-art named entity tagger), then extract the pair as potential spouses.* Eventually, we will explore using polyNER and data programming to extract polymer properties. For instance: *if a sentence contains a polymer name and the words "glass transition", then extract number(s) in the sentence as potential glass transition temperature(s) for that polymer.*

Other approaches use unskilled or semi-expert users to crowdsource the labeling task [15, 7, 38]. Nonetheless, domain expertise is often crucial for identifying and extracting complex scientific entities. Hence, we ask: how can we quickly generate annotated data for scientific named entity recognition?

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

3 Motivation

The complexity of scientific NER is primarily due to the fact that entities, for example biological [12] and chemical [14], can be described in different ways, with vocabularies often specialized to small communities. Such issues arise in the polymer science applications that we focus on here. In principle, International Union of Pure and Applied Chemistry (IUPAC) guidelines define polymer naming conventions [9]. However, such guidelines are not always followed in practice [37]. Polymer names may be reported as source-based names (based on the monomer name), structure-based names (based on the repeat unit), common names (requiring domain-specific knowledge), trade names (based on the manufacturer), and names based on chemical groups within the polymer (requiring context to fully specify the chemistry). Oftentimes, polymers are encoded using acronyms.

These different naming conventions arise in part because a desire for clarity in communications is at odds with the often complicated monomeric structures found in many polymers [1]. For example, sequence-defined polymers, where multiple monomers are chemically bound in a well-defined sequence as in proteins, often defy normal naming practices, as it is not possible to list concisely every monomer and their respective positions [18]. Another class of polymers that often suffer from complicated names are conjugated polymers, which exhibit useful optical and electrical properties. Conjugated polymers are complex due to the co-polymerization of multiple monomers (donor/acceptor units), the type and position of side chains along the polymer backbone, and the coupling between monomer units to control regioregularity [8].

Other challenges arise from the use of labels, structure referents (e.g., "micelles," "nanostructures"), and unusual author-coined acronyms. For example, one author defined the acronym DBGA for N,N-dibenzylglycidylamine and then used the string poly(DBGA) to represent poly(N,N-dibenzylglycidylamine). More naming variations result from typographical variants (e.g., alternative uses of hyphens, brackets, spacing) and alternative component orders.

These issues, which make identifying polymeric names a non-trivial exercise not only for computers but also for experts, arise in many fields with specialized vocabularies. Our long-term goal is to build a hybrid human-computer system in which we leverage both human and machine capabilities for the efficient extraction from text of properties associated with specialized vocabularies. In this work, we focus on the task of identifying polymer names.

4 Design and Implementation

As noted in the introduction, previous approaches to scientific NER have relied on large expert-labeled corpora to train NER tools. Our goal in polyNER is to slash the cost of NER training for new domains by using bootstrap methods to optimize the effectiveness and impact of minimal expert labeling. As shown in Figure 2, rather than having experts review entire papers to identify entity

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

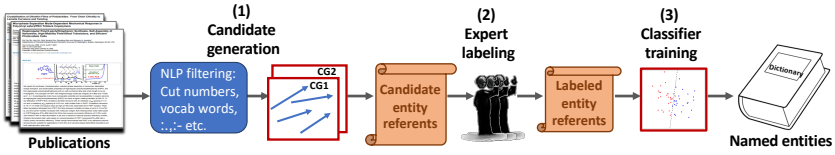


Fig. 1: PolyNER architecture: showing (1) Candidate Generation, which produces candidate named entities from word vectors, (2) Expert Labeling, and (3) Classifier Training, which uses labeled candidates to train supervised ML models for identifying referents.

referents, we use NLP tools to identify a set of promising candidate entity referents (Candidate Generation), then in an Expert Labeling step employ experts to accept or reject those candidates, and finally in a Classifier Training step use the accepted candidates to train an entity classifier

Before turning to the details of the polyNER implementation, we define an NLP filtering process that is used in various places in polyNER to filter out words that are unlikely to be polymer referents. 1) We remove numbers. 2) Hypothesizing that names of scientific entities will not, in general, be English vocabulary words, we remove words found in the SpaCy and NLTK dictionaries of commonly used English words [4, 2]. (We manually remove common polymer names, such as polystyrene and polyethylene, from the dictionaries.) 3) We use SpaCy's part-of-speech tagging functionality to remove non-nouns. 4) We remove unwanted characters (e.g. ':', '.', ',', ':', '-') from the beginning and the end of each candidate, allowing us to recognize, for example, *polyethylene;* (which fails the exact string comparison test against "polyethylene"). 5) We remove plurals (e.g., polyamides, polynorbornenes), as they can represent polymer family names.

4.1 Candidate Generation

This first phase uses word vector representations, vector similarity measures, and minimal domain knowledge to identify a set of high-likelihood ("candidate") entity referents (names, acronyms, synonyms, etc.) in a supplied corpus of full-text documents (in the work presented here, scientific publications).

We first apply the NLP filtering process to reduce false positives. We next face the problem of determining whether a particular string is a polymer referent. String matching only gets us so far: for example, "polyethylene" names a polymer, but "polydispersity" does not. We need also to consider the context in which the string occurs. For example, the polymer name "polystyrene" in a sentence "The melting point of polystyrene is ..." suggests that X may also be a polymer in the sentence "The melting point of X is ...".

NLP researchers have developed a variety of *word embedding* methods for capturing this notion of context. A word embedding method maps each word in a sentence or document to a vector in an n-dimensional real vector space based on the linguistic context in which the word appears. (This mapping may

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

be based, for example, on co-occurrence frequencies of words.) We can then determine the similarity between two words by computing the distance between their corresponding vectors in the feature space. Such vector representations can be created in many different ways [30, 31]. Recently, the efficient neural network-based Word2Vec has become popular [21, 22].

We consider two measures of context-similarity between word vector representations in this step. CG1 uses the Gensim implementation of the Word2Vec algorithm [28] to generate 100-dimension vectors. CG2 employs an alternative FastText word embedding method that considers sub-word information as well as context [3, 11], allowing it to consider word morphology differences, such as prefixes and suffixes. Sub-word information is especially useful for words for which context information is lacking, as words can still be compared to morphologically-similar existing words. We set the length of the sub-word used for comparison—FastText's *n_gram* parameter—to five characters, based on our intuition that many polymers begin with the prefixes "*poly*" or "*poly(*." FastText produces 120-dimension vectors. Both CG1 and CG2 employ the continuous bag-of-words (CBOW) word embedding, in which a vector representation is generated for each word from an adjustable window of surrounding context words, in any order.

We compute a CG1 (Gensim) vector and a CG2 (FastText) vector for each NLP-filtered word in the input corpus, and also for a small set of representative polymer referents. Here we use polystyrene and its common acronym, PS, based on the assumption that polystyrene, as the most commonly mentioned polymer, provides a large number of example sentences in which polymers are mentioned. We can then determine, for each NLP-filtered word, the extent to which it occurs in a similar context to the representative polymers, by computing the similarities between the word's CG1 and CG2 vectors and those for polystyrene and PS. We discard the lower score for each of CG1 and CG2 to obtain two scores per word.

Having thus obtained scores, we then select as candidates, for each of CG1 and CG2, the N highest-scored words, with N selected based on the time available for experts. We also use a rule-based synonym finder to identify synonyms of generated polymer candidates [33]. For example, if polypropylene has been identified as a candidate, then the expression "*polypropylene (PP)*" leads to PP being added to the candidate list.

4.2 Expert Labeling

The previous step produces a set of candidate polymer referents: NLP-filtered strings that have been determined to occur in similar contexts to our representatives. We next employ an expert polymer scientist to indicate, for each such candidate, whether or not it is in fact a polymer referent. The expert simply approves or rejects each candidate via a simple web interface: a task that is more efficient than reading and annotating words in text. The interface (see Figure 2) provides the expert with example sentences as context for ambiguous candidates, and allows the expert to access the publication(s) in which a particular candidate appears when desired.

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

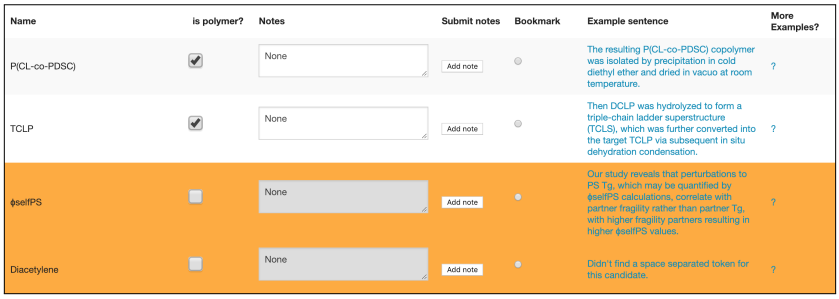


Fig. 2: PolyNER web interface showing annotated candidates. Clicking on "?" delivers up to 25 more example sentences.

4.3 Candidate Discrimination

We next use the expert-labeled data to create an entity classifier. Many classification methods could be applied; we consider three in the work reported here: K Nearest Neighbor (KNN), Support Vector Classifier (SVC), and Random Forest (RF). Previous work has shown that KNNs perform reasonably well in text classification tasks [42]. SVC, an implementation of Support Vector Machines (SVMs), maps data into a feature space in which it can separate the data into two or more sets. RF groups decision trees ("weak learners") to form a strong learner; it produces models that are inspectable, and includes a picture of the most important features. In each case, we use the 100 (Gensim) + 120 (FastText) = 220 dimensions of the two word vectors as input features.

Given limited training and testing data, we evaluate all three classifiers, as implemented within scikit-learn [24], in Section 5.3. We envision that with more annotated data, we will be able to use neural-network-based classifiers.

5 Evaluation

We report on studies in which we evaluate the performance of both the unsupervised Candidate Generation step and various classifiers trained on the labeled data that results from the Candidate Generation and Expert Labeling steps.

5.1 Dataset

We work with two disjoint sets of full-text publications in HTML format from the journal *Macromolecules*: P100 comprising 100 documents with 22 664 sentences and 508 391 (36 293 unique) words or "tokens," and P50 comprising 50 documents with 12 148 sentences and 270 514 (22 571 unique) tokens. For later use in evaluation, we engaged six experts to identify one-word polymer names in P100. They find 467 unique one-word polymer names.

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

5.2 Evaluation of Candidate Generation Methods

Recall that the polyNER candidate generation module employs two candidate generation methods, CG1 and CG2, plus a rule-based synonym finder. We evaluate the performance of both the complete polyNER candidate generation module and its CG1 and CG2 submodules by comparing the sets of candidates that they each generate from P100, both against the 467 one-word polymer names identified by experts in P100, and against two other polymer name extraction methods, CDE and CDE+, plus a sixth compound method formed by combining polyNER with CDE+. We use exact string-matching between candidates and expert-identified names (all lower cased). Results are in Table 1. For each, we evaluate extraction accuracy in terms of precision, recall, and F_1 score. Recall is the fraction of actual positives that are labeled correctly and precision the fraction of predicted positives that are labeled correctly; F_1, the harmonic average of precision and recall, reaches its best value at 1 and worst at 0.

The first two methods considered, CDE and CDE+, serve as baselines. CDE is a state-of-the-art Python package that extracts chemical named entities and associated properties and relationships from text [36]. As CDE aims to extract all chemical compounds, not just polymers, it serves only as a demonstration of an alternative approach in the absence of a polymer NER system. Its recall is high at 74.5% but its precision is, as expected, low at 8.7%. CDE+ extends CDE with manually defined polymer identification rules [39] to achieve a higher precision of 42.2% but a slightly decreased recall of 68.3%. These results emphasize the difficulty of automatically recognizing complex entities such as polymers.

Rows 3 and 4 show performance for CG1 and CG2 when employed independently. Recall that polyNER performs NLP filtering before applying CG1 and CG2. The filtering step eliminate all but 6878 of the 36 293 unique tokens in P100. Recall also that CG1 and CG2 each assign a score to each of the 6878 remaining words based on their context-based vector similarities to polystyrene and PS, and select the N highest scoring. In this evaluation, we set N=500. CG2, which takes word morphology into account, achieves higher precision and recall than does CG1 (41.8% vs. 15.6% precision and 44.8% recall vs. 16.7% recall for CG2 and CG1, respectively). CG2 retrieves more words starting with "poly" (67% of the 500 candidates vs. only 4% for CG1) while CG1 retrieves more acronyms (38% of the 500 candidates contained more upper than lower case letters, vs. 23% for CG2). CG1 returns more false positives. While character level information is useful for unseen words, or in this case for words lacking context information, we cannot dismiss the use of CG1. Authors often introduce polymer names and subsequently use acronyms more heavily, especially for long names. The facts that CG1 returns more acronyms and that there is likely more context information about acronyms, suggests that the performance of CG1, albeit lower, is solely based on context information.

Row 5 shows results for the complete polyNER candidate generator: that is, the combined CG1 and CG2 candidates plus their rule-based extracted synonyms. This method achieves 61.2% recall and 26.0% precision, producing an entity-rich set of candidates without any domain-specific rules and without

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

any (tedious, time-consuming, and costly) expert-annotated corpus of polymer names. We are encouraged to observe that polyNER retrieves polymers not extracted by CDE: the combined recall for PolyNER ∪ CDE+ (row 6 in the table) is 81.6%—higher than CDE itself. This result suggests that polyNER's candidate generation module can be used not only to annotate automatically a diverse set of polymers based on context, but also to improve on the results of more sophisticated hybrid rule- and ML-based NER tools.

Figure 3, which shows every word in P100 in FastText vector space, illustrates the challenges and opportunities inherent in differentiating between polymer and non polymer word vectors. The polymer names (in red and green) form two rather diffuse clusters that overlap considerably with non polymers (in blue). Interestingly, the subset of polymer names that are acronyms (the red points) are clearly clustered.

Table 1: Results when polymer candidates are extracted from our test corpus, P100, via different methods. For each, we show true positives, false positives, false negatives, precision, recall, and F-score.

#	**Method**	**Total**	**TP**	**FP**	**FN**	**Precision**	**Recall**	$\mathbf{F_1}$
1	CDE	3994	348	3646	119	8.7%	74.5%	15.6%
2	CDE+	755	319	436	148	42.2%	68.3%	52.2%
3	CG1	500	78	422	389	15.6%	16.7%	16.1%
4	CG2	500	209	291	258	41.8%	44.7%	43.2%
5	PolyNER	1099	286	813	181	26.0%	61.2%	36.5%
6	PolyNER ∪ CDE+	1495	381	1114	86	25.4%	81.6%	38.8%

5.3 Evaluation of Classifier Training Methods

We next evaluate how well classifiers trained on expert-labeled output from the Candidate Generation phase perform when applied to full-text documents. Here, we make use of our second dataset, P50, to train and test our classifiers, and P100 to validate the trained classifiers.

Before training our classifiers, we need a set of expert-labeled candidates. Thus we first apply the CG1 and CG2 methods of Section 4.1 to P50, generating a total of 897 unique candidates: 500 for CG2 and 466 from CG1, of which 69 overlapped. (We do not apply the rule-based synonym finder here.) Then we employ an expert to label as polymer or non-polymer each of those 897 candidates, producing a new dataset that we refer to as P50-labeled. Note that this task is quick work for the expert, as only 897 words need to be evaluated: the total time required was two hours. This expert review identifies 260 (29.0%) of the 897 as polymers.

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

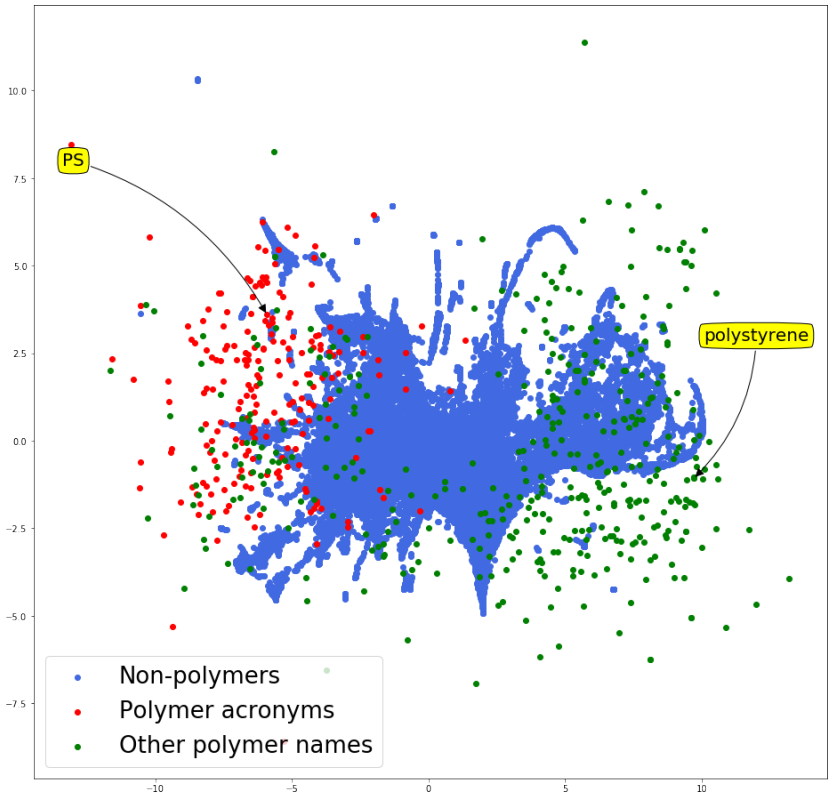


Fig. 3: A two-dimensional representation of all words in P100, generated with the scikit-learn implementation of t-distributed Stochastic Neighbor Embedding (t-SNE) [41]. Of the words identified by experts as polymers, we show acronyms in red and non-acronyms in green; all other words are blue. We label our two representative words. (The t-SNE plot, a dimensionality reduction technique used to graphically simplify large datasets, reduces the 120-dimensional vectors to two-dimensional data points. The axes have no "global" meaning.)

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

Training and validating the classifiers: We next use the 897 expert-labeled words to train our three classifiers. We use 90% (807) for training and hold out 10% (90) for validation.

The left-hand side of Table 2 shows the performance of the different trained classifiers when applied to the P50 hold-out words. Note that performance here is defined with respect to how well the classifier does at predicting the expert labels assigned to the polyNER-generated candidates—not how well the classifier identifies *all* polymer referents in P50, as we do not have the latter information.

All three classifiers obtain between 66.7% and 100.0% precision and between 28.6% and 57.1% recall. SVC achieves the highest recall and F_1 score. The lower part of the table ("combined classifiers") shows that combined classifiers can improve performance. The 3-of-3 method achieves the highest precision (100.0%) but lowest recall, as one might expect. The $\geq$2 method also achieves 100.0% precision but with a higher recall (42.3% vs 28.6%). The $\geq$1 method has the lowest precision but the highest recall at 57.1%.

Table 2: Results when various classifiers (trained on expert-labeled P50 candidates) are applied to P50 holdouts (left) and P100 (right). The results in the bottom two rows are copied from Table 1 for ease of comparison.

Classifier	**Validation on P50 holdouts**			**Testing on P100**		
	Precision	**Recall**	**F_1**	**Precision**	**Recall**	**F_1**
KNN	75.0%	42.8%	54.5%	9.5%	77.8%	16.9%
SVC	66.7%	57.1%	61.5%	16.8%	76.5%	27.5%
RF	100.0%	28.6%	44.4%	51.0%	42.0%	46.1%
Combined						
3-of-3	100.0%	28.6%	44.4%	52.7%	39.1%	44.9%
$\geq$2	100.0%	42.3%	60.0%	22.8%	66.6%	34.0%
$\geq$1	57.1%	57.1%	57.1%	9.1%	90.7%	16.5%
CDE				8.7%	74.5%	15.6%
CDE+				42.2%	68.3%	52.2%

Testing the trained classifiers: We test the trained classifiers by applying each to all 6878 NLP-filtered nouns extracted from P100 and comparing the resulting polymer/non-polymer labels against our ground truth of polymer names extracted from P100 by experts. Results are in the right-hand side of Table 2. RF achieves the highest F_1 score (46.1%) with 51.0% precision and 42.1% recall. While recall is relatively low (fewer entities retrieved), precision is significantly better than that achieved by CDE+. We observe also that combined classifiers can improve precision (52.7%) at the expense of recall, or significantly increase recall (90.1%) at the expense of precision. Users can thus trade off precision and recall, in each case exceeding those achieved by the rule-based CDE+ system.

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

5.4 Discussion

These results are based on only limited training data: just 897 labeled words, of which 260 are polymers. We view the effectiveness of the classifiers trained with these limited data as demonstrating the feasibility of using small amounts of expert-labeled data to bootstrap context-aware word-vector classifiers. Importantly, this whole process was both inexpensive and generalizable to other domains. Candidate generation was fully automated and involved no domain knowledge besides the two representative words, polystyrene and PS. Labeling required just two hours of an expert's time. Classifier training was again automated and involved no domain knowledge.

6 Conclusion

Despite much progress in NLP, scientific named entity recognition (NER) remains a research challenge. A lack of labeled training data in fields such as polymer science limits the use of machine learning models for this task. PolyNER is a generalizable system that can efficiently retrieve and classify scientific named entities. It uses word representations and minimal domain knowledge (a few representative entities) to produce a small set of candidates for expert labeling; labeled candidates are then used to train named entity classifiers.

PolyNER can achieve either 52.7% precision or 90.7% recall when combining classifiers: a 10.5% improvement in precision or 22.4% in recall over a well-performing hybrid NER model (CDE+) that combines a dictionary, expert created rules, and machine learning algorithms. PolyNER's architecture allows users to tradeoff precision and recall by selecting which classifiers are used for discrimination. One out of every four candidates identified by our current polyNER prototype is in fact a polymer. This enrichment relative to the relative paucity of polymers in publications significantly reduces the effort required by experts. Considering that polyNER relies on simple distance from a known polymer(s), and default word embedding parameters, this result is encouraging.

An important issue to explore in future work is whether classifier performance can be improved by providing additional expert-labeled words. We plan to apply active learning [34] to select good candidates. As we generate more expert-labeled candidates, we will explore the use of neural network word vector classifiers to improve accuracy, and the use of polyNER-labeled data to annotate text for other NER approaches, such as bidirectional long short-term memory models. With a view to exploring generalizability, we are also working to apply polyNER to quite different problems, such as extracting dataset names from social science literature. We may explore more recent word representation models, which are pre-trained on large corpora [25, 6].

Acknowledgments

We thank Mark DiTusa, Tengzhou Ma, and Garrett Grocke for contributing manually extracted polymer names. This work was supported in part by NIST

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

contract 60NANB15D077, the Center for Hierarchical Materials Design, and DOE contract DE-AC02-06CH11357, and by computer resources provided by Jetstream [40, 35]. Official contribution of the National Institute of Standards and Technology; not subject to copyright in the US.

References

1. Audus, D. J. & de Pablo, J. J.: Polymer informatics: Opportunities and challenges. ACS Macro Letters **6**(10), 1078–1082 (2017)
2. Bird, S. & Loper, E.: NLTK: The natural language toolkit. In: 42nd Annual Meeting of the Association for Computational Linguistics. p. 31 (2004)
3. Bojanowski, P. et al.: Enriching word vectors with subword information. arXiv:1607.04606 (2016)
4. Choi, J. D. et al.: It depends: Dependency parser comparison using a web-based evaluation tool. In: 53rd Annual Meeting of the ACL. vol. 1, pp. 387–396 (2015)
5. Cohen, A. M. & Hersh, W. R.: A survey of current work in biomedical text mining. Briefings in Bioinformatics **6**(1), 57–71 (2005)
6. Devlin, J. et al.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805 (2018)
7. Gao, H. et al.: Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intelligent Systems **26**(3), 10–14 (2011)
8. Himmelberger, S. & Salleo, A.: Engineering semiconducting polymers for efficient charge transport. MRS Communications **5**(3), 383–395 (2015)
9. Hiorns, R. C. et al.: A brief guide to polymer nomenclature. Polymer **54**(1), 3–4 (2013)
10. Jessop, D. M. et al.: OSCAR4: A flexible architecture for chemical text-mining. Journal of Cheminformatics **3**(1), 41 (2011)
11. Joulin, A. et al.: Bag of tricks for efficient text classification. arXiv:1607.01759 (2016)
12. Kim, J.-D. et al.: Introduction to the bio-entity recognition task at JNLPBA. In: Intl. Joint Workshop on NLP in Biomedicine and its Applications. pp. 70–75 (2004)
13. Krallinger, M. et al.: Overview of the chemical compound and drug name recognition (CHEMDNER) task. In: BioCreative Challenge Evaluation Workshop. vol. 2, p. 2 (2013)
14. Krallinger, M. et al.: CHEMDNER: The drugs and chemical names extraction challenge. Journal of Cheminformatics **7**(1), S1 (2015)
15. Krishna, R. et al.: Visual genome: Connecting language and vision using crowdsourced dense image annotations. Intl. J. Computer Vision **123**(1), 32–73 (2017)
16. Leaman, R. & Gonzalez, G.: BANNER: An executable survey of advances in biomedical named entity recognition. In: Pac. Symp. Bio., pp. 652–663 (2008)
17. Leaman, R. et al.: tmChem: A high performance approach for chemical named entity recognition and normalization. Journal of Cheminformatics **7**(1), S3 (2015)
18. Lutz, J.-F.: Aperiodic copolymers. ACS Macro Letters **3**(10), 1020–1023 (2014)
19. Manning, C. D. et al.: The Stanford CoreNLP natural language processing toolkit. In: ACL (System Demonstrations). pp. 55–60 (2014)
20. Marrero, M. et al.: Named entity recognition: fallacies, challenges and opportunities. Computer Standards & Interfaces **35**(5), 482–489 (2013)
21. Mikolov, T. et al.: Efficient estimation of word representations in vector space. arXiv:1301.3781 (2013)

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

22. Mikolov, T. et al.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Inf. Processing Sys. pp. 3111–3119 (2013)
23. Nadeau, D. & Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)
24. Pedregosa, F. et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
25. Peters, M. E. et al.: Deep contextualized word representations. In: Conf. of the North American Chapter of the Association for Computational Linguistics (2018)
26. Peters, S. E. et al.: A machine reading system for assembling synthetic paleontological databases. PLoS One **9**(12), e113523 (2014)
27. Ratner, A. J. et al.: Data programming: Creating large training sets, quickly. In: Advances in Neural Information Processing Systems. pp. 3567–3575 (2016)
28. Rehurek, R. & Sojka, P.: Software framework for topic modelling with large corpora. In: Workshop on New Challenges for NLP Frameworks (2010)
29. Rocktäschel, T. et al.: ChemSpot: A hybrid system for chemical named entity recognition. Bioinformatics **28**(12), 1633–1640 (2012)
30. Rumelhart, D. E.: Learning internal representations by back-propagating errors. Parallel Distributed Processing **1**, 318–362 (1986)
31. Sabes, P. N. & Jordan, M. I.: Reinforcement learning by probability matching. In: Advances in Neural Information Processing Systems. pp. 1080–1086 (1995)
32. Sang, E. F. T. K. & De Meulder, F.: Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In: 7th Conference on Natural Language Learning. pp. 142–147 (2003)
33. Schwartz, A. S. & Hearst, M. A.: A simple algorithm for identifying abbreviation definitions in biomedical text. In: Pac. Symp. Bio., pp. 451–462 (2002)
34. Settles, B.: Active learning. Synthesis Lectures on Artificial Intelligence and Machine Learning **6**(1), 1–114 (2012)
35. Stewart, C. A. et al.: Jetstream: A self-provisoned, scalable science and engineering cloud environment (2015), `https://doi.org/10.1145/2792745.2792774`
36. Swain, M. C. & Cole, J. M.: ChemDataExtractor: A toolkit for automated extraction of chemical information from the scientific literature. Journal of Chemical Information and Modeling **56**(10), 1894–1904 (2016)
37. Tamames, J. & Valencia, A.: The success (or not) of HUGO nomenclature. Genome Biology **7**(5), 402 (2006)
38. Tchoua, R. B. et al.: A hybrid human-computer approach to the extraction of scientific facts from the literature. Procedia Computer Science **80**, 386–397 (2016)
39. Tchoua, R. B. et al.: Towards a hybrid human-computer scientific information extraction pipeline. In: 13th Intl. Conference on e-Science. pp. 109–118 (2017)
40. Towns, J. et al.: XSEDE: Accelerating scientific discovery. Computing in Science & Engineering **16**(5), 62–74 (2014)
41. van der Maaten, L. J. P. & Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research **9**(Nov), 2579–2605 (2008)
42. Yang, Y. & Liu, X.: A re-examination of text categorization methods. In: 22nd Annual International ACM SIGIR Conference. pp. 42–49. ACM (1999)
43. Zhang, C. et al.: GeoDeepDive: Statistical inference using familiar data-processing languages. In: ACM SIGMOD Conference. pp. 993–996 (2013)

Tchoua, Roselyne; Ajith, Aswathy; Hong, Zhi; Ward, Logan; Chard, Kyle; Audus, Debra; Patel, Shrayesh; de Pablo, Juan; Foster, Ian.
"Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort."
Paper presented at ICCS 2019: The International Conference on Computational Science, Faro, Portugal. June 12, 2019 - June 14, 2019.

2018 IEEE 13th Nanotechnology Materials and Devices Conference (NMDC)

Characterization of Graphene Conductance Using a Microwave Cavity

Jan Obrzut
Materials Science and Eng.
National Institute of *Standards and Technology*
Gaithersburg, Maryland, USA
jan.obrzut@nist.gov

Keith White
Materials Science and Eng.
National Institute of *Standards and Technology*
Gaithersburg, Maryland, USA
keith.white@nist.gov

Yanfei Yang
Quantum Measurement Div.
National Institute of *Standards and Technology*
Gaithersburg, Maryland, USA
yanfei.yang@nist.gov

Randolph E. Elmquist
Quantum Measurement Div.
National Institute of *Standards and Technology*
Gaithersburg, Maryland, USA
randolph.elmquist@nist.gov

***Abstract*— Surface conductance of graphene films is investigated in a non-contact microwave cavity operating at 7.4543 GHz. This cavity technique characterizes specimens with high accuracy, without disruption of morphological integrity and without the processing required for definition and application of electrical contacts. The thickness of the conducting 2D material does not need to be explicitly known. Measurement results are illustrated for epitaxial mono-layer graphene formed on 4H-SiC(0001) wafers by annealing the substrate via high fidelity Si sublimation. We show that the resonant microwave cavity is sensitive to the surface conductance of atomically thin nano-carbon films. The method is applied to characterize diffusion of water through a thin film moisture barrier coated over epitaxial monolayer graphene. The monolayer graphene is lightly *n*-type doped and therefore the resulting surface resistance value becomes sensitive to complexation with *p*-type molecular dopants such as water in moisturized air. Our results suggest that coating graphene surface with 100 nm thick Parylene-C film serves as effective moisture barrier. The observed change in conductance of Parylene-C specimens after environmental stress is within 6 % of the pre-test value and would satisfy the typical reliability requirements.**

Keywords—microwave cavity, surface conductance, epitaxial graphene, moisture barrier, molecular doping.

I. Introduction

Recent years have witnessed many breakthroughs in research as well as a significant investment in the advancement of the mass production of graphene materials. The market of graphene applications is essentially driven by progress in the production of graphene with properties appropriate for the specific application. There are several methods being used and developed to prepare graphene-like materials of various dimensions, morphology and number of layers. However, they are being offered with unreferenced quality, defect density and type, as these parameters affect these material's fundamental properties and behaviors. In this context, a graphene reference material for which the surface conductance can be related to number of layers, mobility and fundamental physical constants can be a breakthrough that may change the face of the industry. Here we characterize conductance of a mono layer epitaxial graphene on silicon carbide, where mobility and density of charge carriers are traceable through Hall resistance measurements. Surface conductance is determined at room temperature by a non-contact microwave cavity measurement. Surface conductance is determined at room temperature by a non-contact microwave cavity measurement. The method is nondestructive, experimentally simple and it does not require any electrical contacts that would otherwise obscure complexation with molecular dopants. The monolayer epitaxial graphene in conjunction with the non-contact microwave cavity method has the potential to be used as a 2D surface conductance/resistance reference material that would enable the industry to assess the quality and the corresponding electronic properties of their product. We investigate affinity of graphene to molecular complexation with environmental water under environmental stress conditions.

II. Experimental

A. Preparation of epitaxial graphene

Mono-layer epitaxial graphene on the Si-terminated face (SiC/G) was grown on high purity 4H-SiC(0001) 330 μm thick wafers. The substrates were annealed at 1900°C in Argon at (101-105) kPa using a controlled Si sublimation process, which stops at one mono-layer on Si-terminated face [1]. The wafers were then diced to obtain specimen size of 7.8 mm × 3.7 mm. After removing the graphene material on the C-terminated face, several SiC/G samples were tested directly in the microwave cavity on the native substrate. Hall resistance (ρ_{xy}), longitudinal resistance (ρ_{xx}), carrier concentration (n) and Hall mobility (μ) were measured on the corresponding samples at temperatures from 1.5 K to 30 K, using a Hall-bar device having a channel size 100 μm × 600 μm. At temperature of 298 K the Hall mobility of SiC/G is about 3200 $cm^2V^{-1}s^{-1}$, and the measured σ_G corresponds to the density of charge carriers, n, of about 1.7×10^{11} cm^{-2}. Due to interaction of the monolayer graphene with impurity scattering charge carriers at the graphene SiC substrate interface, the monolayer graphene is slightly *n*-type doped and therefore the resulting surface resistance value becomes sensitive to complexation with *p*-type molecular dopants such as water in moisturized air [2-4]. We applied Parylene-C thin film conformal coating as a protective barrier. Parylene-C

Obrzut, Jan; White, Keith; Yang, Yanfei; Elmquist, Randolph.
"Characterization of graphene conductance using a microwave cavity."
Paper presented at IEEE 13th Nanotechnology Materials & Devices Conference (NMDC 2018), Portland, OR, United States. October 14, 2018 - October 17, 2018.

(poly-chloroxylylene) was deposited on the surface of selected SiC/G specimens in vacuum by pyrolysis at 600 ºC, followed by polymerization at 25 ºC, to obtain coatings about 100 nm thick [3]. Permeability of Parylene coatings was evaluated by exposing specimens to humid air and measuring a decrease of SiC/G surface conductance in correlation with the number of charge carriers immobilized by complexation with water molecules.

B. *Measurement of surface conductance by the non contact cavity perturbation method*

In our earlier work [5, 6] we showed that for a small conducting specimen inside a rectangular cavity operating in the TE_{10n} mode, solution of the imaginary part of cavity perturbation equation can be expressed by (1):

$$y'' = \varepsilon'' 4x - 2b'' \tag{1}$$

where, $x = V_s / V_0$, $y'' = 1/Q_s - 1/Q_0$, V_0 is the volume of the cavity, V_s is the volume of the specimen ($V_0 >> V_s$). Q_0 is the quality factor of the empty cavity and Q_s is the quality factor of the cavity loaded with the specimen. In the case of conducting materials with high mobility of delocalized charge carriers, such as graphene, the energy loss resulting from dipolar polarization and polarization of bound charges can be neglected, and the dielectric loss factor, ε''_r, is fundamentally equivalent to conductivity. This relation can be expressed in terms of surface conductance [5], which after substituting into (1) leads to (2):

$$\frac{1}{Q_x} - \frac{1}{Q_0} = \sigma_G \frac{2wh_x}{\pi \varepsilon_0 f_x V_0} c_\omega \tag{2}$$

where, σ_G is the surface conductance of graphene, ε_0 is the permittivity of free space, wh_x is the area of the graphene layer in the cavity, Q_x is the quality factor of cavity partially loaded with the graphene at wh_x, $c_\omega = (f_x/f_0)^2$ accounts for frequency correction due to substrate polarization, f_x is the resonant frequency of sample-loaded cavity and Q_0, and V_0 are the same as defined in (1). The quality factor is obtained from the resonant peak according to the conventional half power bandwidth formula as $Q_x = f_x/\Delta f_x$. Equation (2), from which we determine σ_G, does not depend explicitly on the thickness of the graphene layer, but on its area, wh_x, which can be determined with much higher accuracy than t_G. In the range of h_x where b'' is constant, (2) reduces to a linear equation with slope of σ_G. In the case when the surface conductance is comparable with that of the substrate and cannot be neglected, the conductance of the graphene layer can be extracted using a parallel admittance model of combined conductance described in Ref. [6]. Our cavity test fixture design shown in Fig. 1 employs a WR 90 waveguide (a=10.16 mm, b=22.86 mm, l_z = 127.0 mm) operating in the microwave frequency range of 6.7 GHz to 13 GHz. The fixture is connected to a network analyzer (Agilent N5225A) with semi-rigid coaxial cables and coaxial to WR 90 coupling adapters. The walls of the cavity are implemented via WR 90 couplers, which are near cross-polarized in respect to the waveguide polarization to achieve optimal power loading into cavity while maximizing the quality factor. The specimen is inserted into the cavity through a slot machined in the center of the cavity, where the electric field, $\boldsymbol{E}_x$, attains a maximum value. The specimen insertion (h_x), and the corresponding area of the material in the cavity (wh_x) are controlled by a stage.

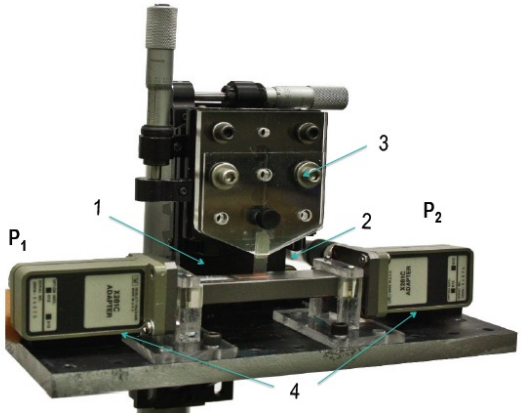


Fig. 1. Microwave cavity test fixture. (1) WR90 waveguide, (2) Specimen, (30) Specimen holder, (4) Coax to WR90 cross-polarized couplers, ports to vector network analyzer ($P_1$and P_2).

During measurements, the specimen is partially inserted in steps, while the scattering parameter, $\mathbf{S}_{21}$, is recorded. The implemented non-contact microwave cavity technique is applicable to materials having surface conductance from 10^{-6} S to 10 S. The surface conductance can be determined with the combined relative uncertainty of 1.5 %.

Results and Discussion

Fig. 2 exemplifies the scattering parameter |magnitude, $|\mathbf{S}_{21}|$, of the TE_{103} mode measured for the epitaxial graphene as a function of the specimen insertion into cavity. The height of the resonant peaks of the SiC/G decrease and widen considerably with increasing specimen insertion due to conductance of the graphene layer. In comparison, the height and width of the resonant peaks of SiC (not shown) remains relatively unchanged with increasing specimen insertion, indicating low conductivity. From (2), σ_s of SiC substrate is in the range of 10^{-7} S. When aligned, the position of resonant peaks of SiC/G and the SiC substrate overlaps. Thus, the frequency shift of the resonant peaks in Fig 2 to lower frequencies is primarily due to the dielectric permittivity of the SiC substrate for which we find the dielectric constant, ε'_r of SiC to be 9.4 at the operational frequency of 7.435 GHz. Therefore, the frequency shift due to graphene is zero, that is, the real part of the relative permittivity of our epitaxial mono-layer graphene $\varepsilon'_r \approx 1.0$, which agrees well with the dielectric permittivity of truly 2D materials. This indicates that our SiC/G is of very high quality 2D monolayer graphene, practically free from polarizable defects. Evidently, we observe a coherent-like

Obrzut, Jan; White, Keith; Yang, Yanfei; Elmquist, Randolph.
"Characterization of graphene conductance using a microwave cavity."
Paper presented at IEEE 13th Nanotechnology Materials & Devices Conference (NMDC 2018), Portland, OR, United States. October 14, 2018 - October 17, 2018.

response from a 2D lattice composed of C-2p_z conjugated electronic system with high charge carriers mobility.

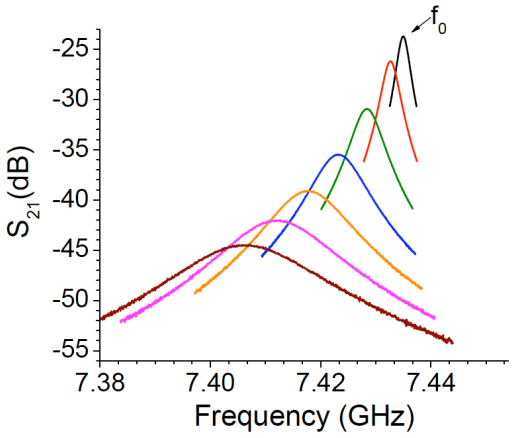


Fig. 2 Scattering parameter $|\mathbf{S}_{21}|$ of the resonant peak TE_{103} measured for SiC/G specimen. The peak position moves from f_0 to lower frequencies with increasing specimen insertion area (wh_x) due to dielectric constant od SiC. The peak decreases and widens with increasing insertion due to conductance of the graphene layer; f_0 = 7.434935 GHz.

In contrast, in the case of graphene materials with multi-layer graphene patches, voids, 3D or amorphous carbon impurities, polarization at the grain boundaries gives rise to di-polar polarization causing an additional shift in resonant frequency, beyond that corresponding to polarization from the dielectric substrate.

Fig. 3 shows graphical representation of (2), with experimental Q_x and h_x data for unprotected SiC/G at several RH conditions. These are compared with specimens coated with 100 nm thick barrier layer of Parylene-C. The solid lines are linear fits to (2) through the data points. In the presented notation, the surface conductance, σ_G, is obtained directly from the slope of the linear portion of the plots. When exposed to environmental moisture, surface conductance of freshly prepared SiC/G drops to certain saturated level. The effect of molecular complexation of water on SiC/G conductance is entirely reversible. After heating the exposed specimen at 130 ºC for few hours under Argon-moisture-free atmosphere, conductance increases back to its initial value, $\sigma_G = 8.7507\times10^{-5}$ S (see Fig. 3 plot 1).

From the slope of plots 2-4 in Fig 3 we find that at the relative humidity RH=20 % the 24 h saturated value of σ_G decreases to 5.593×10^{-5} S, at RH=30 % σ_G decreases to 5.264×10^{-5} S and at RH of 60 % the 24 h saturated value of σ_G approaches 4.5525×10^{-5} S. Based on our earlier investigations we may assume that within our experimental conditions the charge carriers mobility, μ_0 = 3200 $cm^2V^{-1}s^{-1}$, remains negligibly affected by the complexation process [7]. Therefore, density of charge carriers is directly relatet to conductance, $n_i = \sigma_{G\text{-}0}/e\,\mu_0$, and it decreases from its initial value n_1= 17.09 $\times10^{10}$ cm^{-2} to n_2=10.924 $\times10^{10}$ cm^{-2} at RH 20%, n_3=10.028 $\times10^{10}$ cm^{-2} at RH 30% and to n_4=8.891 $\times10^{10}$ cm^{-2} at RH 60%. In comparison, conductance of the specimens coated with Parylene protective layer remains essentially unchanged when exposed to ambient humidity-temperature (50% RH 20 ºC) for several weeks. Yet, after exposure to an accelerated environmental stress we noticed a decrease in surface conductance. Plots 5 and 6 in Fig. 3 show measurement results obtained for Parylene coated specimens that were exposed to accelerated environmental stress at 60 ºC, 85 % RH for 336 h (two weeks).

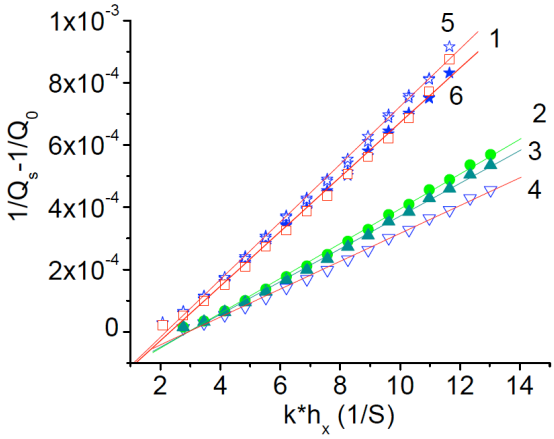


Fig.3. Plots of $(1/Q_s\text{-}1/Q_0)$ vs normalized specimen area (kh_x) for monolayer epitaxial graphene: (1-open squares) after heating in argon, (2-filed circles) exposured RH 20%, (3-filed triangles) RH 30%, (4-open triangles) RH 60%, (5-open stars) coated with Parylene and (6-filled stars) Parylene coated after environmental stress. The solid lines are linear fits to (2) through the data points where the slope respresents conductance, σ_G.

The observed decrease in conductance from that of Argon conditioning level of 9.27×10^{-5} S (plot 5) to 8.7507×10^{-5} S after the stress (plot 6), is detectable by our non-contact measurement method. The corresponding change in the number of charge carriers $\Delta n \approx 1.01 \times10^{10}$. The observed change in conductance is relatively small, about 6% and would likely satisfy the typical reliability requirements for SiC/G as a conductance reference for other graphene materials

Assuming that complexation takes one electron per water molecules then about 1 $\times10^{10}$ water molecules can permeate through 1 cm^2 area of 100 nm thick barrier over 336 h diffusion time. Transferring these numbers from the molecular nanoscale to technical system of units (kg, m, s), one can find that (Δn/6.07 $\times10^{23}$ mol) $\times18$ $\times10^{-3}$ kg/mol) $\approx 3 \times10^{-16}$ kg of water permeates over 1×10^{-4} m^2 area , distance 10^{-7} m in 1.21 $\times$ 10^6 s. These results are significant and prove the usefulness of the presented non-contact measurement methodology to evaluate permeability of nano-size thin film coatings in general.

III. Summary

We characterized conductance of a mono layer epitaxial graphene on silicon carbide with mobility and density of charge carriers traceable through Hall resistance. Surface conductance is determined at room temperature through a non-contact resonant microwave cavity measurement, and since no additional processing is required, it preserves the integrity of the conductive graphene layer. Like other non-contact methods (i.e., ellipsometry and optical density) it allows characterization

Obrzut, Jan; White, Keith; Yang, Yanfei; Elmquist, Randolph.
"Characterization of graphene conductance using a microwave cavity."
Paper presented at IEEE 13th Nanotechnology Materials & Devices Conference (NMDC 2018), Portland, OR, United States. October 14, 2018 - October 17, 2018.

with high speed and efficiency, compared to transport measurements where sample contacts must be defined and applied in multiple processing steps. By using 100 nm thick layer of Parylene-C conformal coating as a protective encapsulation layer, stability of graphene electrical properties is considerably improved. After exposure to accelerated stress at 60 ºC, 85 % RH for 336 h the observed change in surface conductance was less than 6 % of its pre-stress test value, which would satisfy the typical reliability requirements for consumer electronics. We believe that such epitaxial graphene can be used as 2D conductance/resistance reference material that would enable the industry to assess the quality and the corresponding electronic properties of their product, without ambiguity. The presented results can be applied for permeability characterization of nanoscale aqueous barriers. Our approach allows determining diffusion of water molecules through thin protective encapsulation layers coated on graphene surface with resolution unobtainable by conventional methods.

ACKNOWLEDGMENT

K.W. thanks for the 2016 NIST Summer Undergraduate Fellowship support.

DISCLAIMER

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

[1] Y. Yang, L.-I. Huang, Y. Fukuyama, F.-H. Liu, M. A. Real, P. Barbara, C.-T. Liang, D. B. Newell, and R. E. Elmquist, “Low carrier density epitaxial graphene devices on SiC,” Small vol. 11 pp. 90-95, 2015.

[2] Y. Yang, K. Brenner, and R. Murali, “The influence of atmosphere on electrical transport in graphene,” Carbon, vol 50, pp. 1727-11733, 2012.

[3] A. F. Rigosi, C.I. Liu, Bi Yi W, H.Y. Lee, M. Kruskopf, Y. Yang, H. M. Hill, J. Hu, E. G. Bittle, J. Obrzut, A. R. Hight Walker, R. E. Elmquist, D. B. Newell, “Examining epitaxial graphene surface conductivity and quantum Hall device stability with Parylene passivation”, Microelectronics Eng., vol. 194, pp. 51-55, 2018.

[4] C. Melios, C. E. Giusca, V. Panchal and O. Kazakova, “Water on graphene: review of recent progress,” 2D Mater., vol 5, p. 022001, 2018

[5] J. Obrzut, C. Emiroglu, O. Kirillov, Y. Yang, R.E. Elmquist, “ Surface conductance of graphene from non-contact resonant cavity,” Measurement, vol. 87, pp. 146–151, 2016.

[6] N.D. Orloff, J. Obrzut, C.J. Long, T. Lam, P. Kabos, D.R. Novotny, J.C. Booth, and J.A. Liddle, “Dielectric characterization by microwave cavity perturbation corrected for nonuniform fields,” IEEE Trans. Microw. Theory Techn., vol. 62, pp. 2149-2159, 2014.

[7] C. Chuang, D. Chiba, Y. Yang, S. Pookpanratana, C. Hacker, C. T. Liang, R.E. Elmquist,“Chemical-doping-driven crossover from graphene to ordinary metal in epitaxial graphene grown on SiC”, Nanoscale, 2017, 9, 11537.

T

Obrzut, Jan; White, Keith; Yang, Yanfei; Elmquist, Randolph.
”Characterization of graphene conductance using a microwave cavity.”
Paper presented at IEEE 13th Nanotechnology Materials & Devices Conference (NMDC 2018), Portland, OR, United States. October 14, 2018 - October 17, 2018.

Proceedings

MEMS Non-Absorbing Electromagnetic Power Sensor Employing the Effect of Radiation Pressure †

Ivan Ryger [1,2,*], Alexandra Artusio-Glimpse [3], Paul Williams [3], Gordon Shaw [4], Matthew Simons [3], Christopher Holloway [3] and John Lehman [3]

[1] Associate of the National Institute of Standards and Technology, Boulder, CO 80305, USA
[2] Department of Physics, University of Colorado, Boulder, CO 80309, USA
[3] National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA; alexandra.artusio-glimpse@nist.gov (A.A.-G.); paul.williams@nist.gov (P.W.); matthew.simons@nist.gov (M.S.); christopher.holloway@nist.gov (C.H.); john.lehman@nist.gov (J.L.)
[4] National Institute of Standards and Technology, 100 Bureau Dr, Gaithersburg, MD 20899, USA; gordon.shaw@nist.gov
* Correspondence: ivan.ryger@nist.gov; Tel.: +1-303-497-5795
† Presented at the Eurosensors 2018 Conference, Graz, Austria, 9–12 September 2018.

Published: 24 December 2018

Abstract: We demonstrate a compact electromagnetic power sensor based on force effects of electromagnetic radiation onto a highly reflective mirror surface. Unlike the conventional power measurement approach, the photons are not absorbed and can be further used in the investigated system. In addition, the exerted force is frequency-independent, yielding a wide measurement frequency span being practically limited by the wavelength-dependent mirror reflection coefficient. The mechanical arrangement of two sensing elements in tandem suppresses the influence of gravity and vibrations on the power reading. We achieve a noise floor of about 1 W/√Hz and speed of 100 ms, being practically limited by sensor's dynamics and lock-in amplifier filter settling time.

Keywords: radiation pressure; capacitive sensor; lock-in amplifier; silicon micromachining; acoustic noise suppression; tilt immunity; distributed bragg reflector; Archimedian spiral spring

1. Introduction

The industrial use of electromagnetic power sources (e.g., laser-based additive manufacturing, and microwave plasma processes in nanotechnology) is constantly rising. Accurate knowledge of delivered power is vital for understanding the underlying physical processes and simultaneously ensuring consistent process quality. The most accurate primary standards for absolute SI-traceable measurement of high electromagnetic (EM) power at National Institute of Standards and Technology are based on comparison of heat generated by absorption of the EM energy to its SI-traceable Joule electrical equivalent [1]. Calorimeters used for this purpose require a total absorption of electromagnetic energy in a black body rendering it unavailable for further use. They are from their physical nature bulky and slow instruments integrating out short-term fluctuations of the measured power. On the other hand, often industrial EM power sources are measured by a small portion of the energy being picked off by a power splitter. However, due to fluctuations of the splitting ratio, the measurement error is oftentimes at the level of measurand [2]. For this reason, there exists an open gap between high precision primary power standards for EM power and compact and fast industrial sensors (often photodetectors in laser applications). Measuring power by means of radiation pressure is attractive because it stems from the linear and frequency-independent relationship between the incident laser power and the force it exerts on a mirror (1)

Ryger, Ivan; Artusio-Glimpse, Alexandra; Williams, Paul; Shaw, Gordon; Simons, Matthew; Holloway, Christopher; Lehman, John.
"MEMS non-absorbing electromagnetic power sensor employing the effect of radiation pressure."
Paper presented at Eurosensors 2018, Graz, Austria. September 9, 2018 - September 12, 2018.

$$\boldsymbol{F = 2Pr\,cos(\theta)/c} \quad (1)$$

where F, P, θ and c are force applied onto mirror, optical power, angle of incidence towards plane normal and speed of the light, respectively. The reflectivity dependent constant $r = R + (1 - R)\,\alpha/2$ accounts for light momentum transfer in the case of finite reflectivity R and absorptivity α of the mirror [3]. The measurement of the force effects of EM radiation has been carried out successfully in many publications [4–8]. However, these laboratory instruments were found to be delicate and sensitive to environmental conditions and thus impractical for routine operation. But now, owing to the availability of high-sensitivity industrial scales and dielectric mirrors that can be optimized for very high reflectivity (typically >99.99% for optical and >90% for RF), a SI-traceable high optical power measurement by means of the force exerted by photons onto a highly reflective mirror has been demonstrated [3]. This principle allows an energy-carrying beam to be bounced off a measurement mirror and still used in the target application. However, this concept suffers from an inherent sensitivity to inertial forces (acoustic vibrations and tilt-dependent projection of gravity). As a proposed solution, we demonstrate a miniaturized version of a radiation pressure sensor that could potentially bridge the gap between the accurate primary EM power standards and better-performing but less accurate industrial detectors. We describe here a design to inherently suppress the environmental effects on the sensor reading, thus approaching the needs of manufacturing floor. Its compact size and ability of direct force calibration are enabling features for its use inside sources of electromagnetic energy, making a considerable paradigm shift from calibrated detectors towards calibrated energy sources. We demonstrate the performance of this sensor to measure microwave power at a frequency of 15 GHz.

2. Sensor Design and Characterization

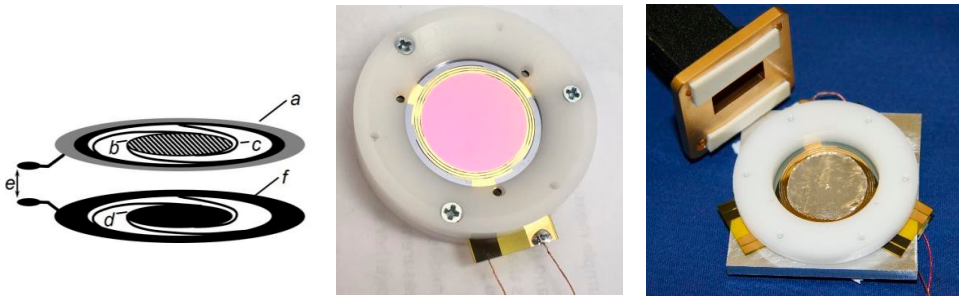
The sensor consists of a silicon micromachined parallel plate capacitor, both of whose disk electrodes (diameter 20 mm) are attached to a rigid frame through weak springs (width 465 μm, thickness 380 μm, length 45 mm) in the shape of an Archimedean spiral (Figure 1a). For the optical sensor, a highly reflective dielectric mirror (optimized for reflectance with a value of 0.9992 ± 0.0002 at 1.07 μm wavelength and 45° angle of incidence) is deposited onto the front side of one disk electrode. The rear side of the silicon chip is covered by continuous Ti/Au metallization to prevent photoconductive effects from affecting the capacitance reading. An additional optimized Ti/Au layer is deposited on top of the Archimedian springs to compensate the bimorph thermal bending of springs. Two identical chips are separated by thin (~42.5 μm) insulating polyimide foil with metallized electrical contacts (Figure 1b shows such a device for laser power measurement) and clamped together. In the case of a microwave power sensor, Figure 1c shows a setup with a front mirror made from an aluminum foil approximately 30 μm thick. Due to our lack of available sensor chips at the time, for this microwave sensor, the bottom element of the sensor was replaced by a machined aluminum plate (Figure 1c).

To obtain a calibration between a measured electrical signal and the applied optical power, an accurate knowledge of electro-mechanical parameters (spring constant, capacitor plate spacing and capacitance gradient with respect to plate spacing) of our transducer is needed. For this purpose, we developed an opto-electronic characterization technique, where we apply a time-varying electrostatic bias voltage to the capacitor's electrodes and measure the axial displacement of the mirrors [9]. Simultaneously we record the geometrical capacitance measured by a LCR meter as a function of capacitor plate spacing (Figure 2a), yielding the knowledge of the capacitance gradient. We employ a dual axis heterodyne interferometer to measure the displacements of both front and back mirror faces. The optical setup comprises two single pass interferometers in a cat's-eye arrangement [10], that have been proven to be relatively insensitive to tilt errors. Using these measurements, we determine the initial spacing of the capacitor's electrodes, the mechanically inactive (parasitic) capacitance and the spring constant.

To demonstrate the benefits of dual spring tandem arrangement we mounted the sensors from Figure 1b,c onto rotational stage and recorded the sensor's geometrical capacitance as a function of

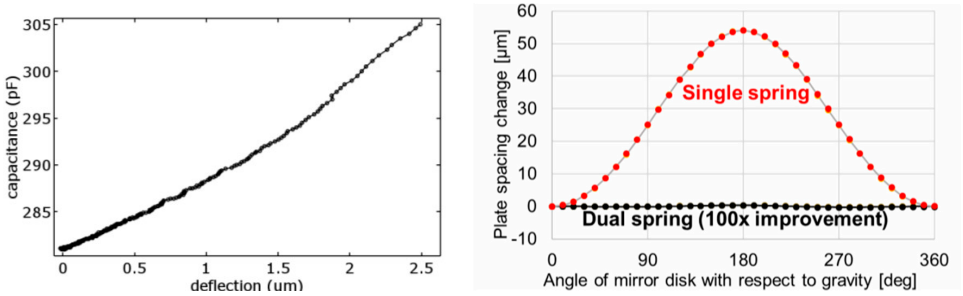
Ryger, Ivan; Artusio-Glimpse, Alexandra; Williams, Paul; Shaw, Gordon; Simons, Matthew; Holloway, Christopher; Lehman, John.
"MEMS non-absorbing electromagnetic power sensor employing the effect of radiation pressure."
Paper presented at Eurosensors 2018, Graz, Austria. September 9, 2018 - September 12, 2018.

tilt angle. Then this value was mapped into equivalent plate spacing based on known capacitance calibration (Figure 2a). We measure suppression of tilt-dependent gravity projection by factor of 100 in the case of two identical silicon chips operating in tandem compared to a device with second electrode being a rigid aluminum plate.



(**a**) (**b**) (**c**)

Figure 1. (**a**) Schematic view of the dual spring radiation pressure sensor. It consists of two disk electrodes, *a* with mirror on top *b* and a metallic coating *d* on the back side. These moving disks are attached to the supporting annulus *a* through three Archimedian springs *c*. Electrical measurement is provided through contacts *e*. (**b**) Photograph of fabricated dual spring optical sensor. (**c**) Microwave power sensor with bottom element replaced by an aluminum plate, next to the WR62 waveguide.



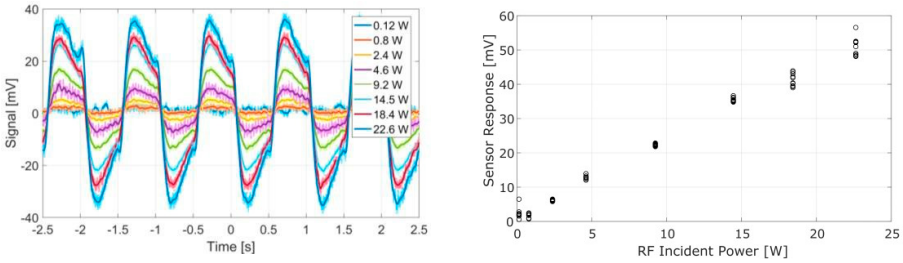
(**a**) (**b**)

Figure 2. (**a**) Plot of the sensor's geometrical capacitance as a function of deflection. This is to determine the capacitance gradient, parasitic capacitance and initial plate spacing (comparison of two methods). (**b**) Plot of capacitor plate spacing as a function of the tilt angle. Zero angle corresponds to moveable mirror surface facing upwards and fixed plate downwards, 180° angle corresponds to fixed plate facing upwards and moveable electrode facing downwards.

3. Microwave Power Experiment

Due to the lack of available chips during the test period we used a single spring device with machined aluminum backplate instead of dual spring device. The sensor was mounted directly onto the output flange of WR62 waveguide. The output bridge signal (proportional to plate capacitance) was fed to a dual phase lock-in amplifier and its output was recorded by a digital storage oscilloscope. The microwave power at 15 GHz was provided by a signal generator connected to a travelling wave tube amplifier. Two waveguide ferrite isolators were placed in the signal path to dissipate the power reflected from the sensor. The output power was monitored with a directional coupler which picked off a fraction of the microwave power reflecting from the mirror, further attenuated it and measured it with a bolometer power head. The signal generator was modulated by a square wave of 1 Hz frequency and 50% duty cycle. To decrease the contribution of the un-correlated noise we triggered the oscilloscope from the same modulation source and used waveform averaging [9]. The experiment was performed in a Faraday cage padded with foam microwave absorbers. To monitor the drift of the sensor, we measured its temperature during the experiment by an attached thermocouple

thermometer with a readout outside the measurement chamber. A plot of the transient response of the sensor is shown in Figure 3a. The linearity of the measurement is evaluated by plotting the sensor response as a function of power in Figure 3b.



(**a**) (**b**)

Figure 3. (**a**) Transient response of the miniature radiation pressure power meter to a microwave power. The slow creep of the sensor baseline is attributed to a high pass filter at the input of the oscilloscope (AC coupling). (**b**) Plot of the sensor's response peak signal against varying incident power; following the linear trend.

4. Conclusions

In this contribution we demonstrated the manufacturing process and characterization of a MEMS power sensor for detecting the force effects of EM radiation. We demonstrated its performance by measuring the output power of microwave frequency radiation. We show an achieved noise limit of about 1W/√Hz and response time around 100 ms that is limited by sensor mechanics and the lock-in amplifier's filter settings. We also demonstrate benefits of tandem spring arrangement, suppressing the tilt-dependent projection of gravity force by factor of 100. Owing to its small mechanical dimensions, noise-rejecting dual spring topology and optimized design, this device extends the applicable measurement range towards lower power levels and broad frequency range.

References

1. Williams, P.A.; Hadler, J.A.; Cromer, C.; West, J.; Li, X.; Lehman, J.H. Flowing-water optical power meter for primary-standard, multi-kilowatt laser power measurements. *Metrologia* **2018**, *55*, 427–436.
2. Spidell, M.; Hadler, J.A.; Stephens, M.; Williams, P.; Lehman, J.H. Geometric contributions to chopper wheel optical attenuation and uncertainty. *Metrologia* **2017**, *54*, L19–L25.
3. Williams, P.; Hadler, J.A.; Maring, F.C.; Lee, R.; Rogers, K.A.; Simonds, B.J.; Spidell, M.T.; Feldman, A.D.; Lehman, J.H. Portable, high-accuracy, non-absorbing laser power measurement at kilowatt levels by means of radiation pressure. *Opt. Express* **2017**, *25*, 4382–4392.
4. Nichols, E.F.; Hull, G.F. A7 preliminary communication on the pressure of heat and light radiation. *Phys. Rev. Ser. I* **1901**, *13*, 307–320.
5. Lebedev, P.N. Experimental examination of light pressure. *Ann. Phys.* **1901**, *6*, 1–26.
6. Cullen, A.L. Absolute power measurement at microwave frequencies. *Proc. IEEE* **1952**, *92*, 100–111.
7. Ma, D.; Garrett, J.L.; Munday, J.N. Quantitative measurement of radiation pressure on microcantilever in ambient environmnent. *Appl. Phys. Lett.* **2015**, *106*, 091107.
8. Agatsuma, K.; Friedrich, D.; Ballmer, S.; DeSalvo, G.; Sakata, S. Nishida, E. Kawamura, S. Precise measurement of laser power using an optomechanical system. *Opt. Express* **2014**, *22*, 2013–2030.

Ryger, Ivan; Artusio-Glimpse, Alexandra; Williams, Paul; Shaw, Gordon; Simons, Matthew; Holloway, Christopher; Lehman, John.
"MEMS non-absorbing electromagnetic power sensor employing the effect of radiation pressure."
Paper presented at Eurosensors 2018, Graz, Austria. September 9, 2018 - September 12, 2018.

9. Ryger, I.; Artusio-Glimpse, A.B.; Williams, P.; Tomlin, N.; Stephens, M.; Rogers, K.; Spidell, M.; Lehman, J., Micromachined force scale for optical power measurement by radiation pressure sensing. *IEEE Sens. J.* **2018**, *18*, 7941–7948.
10. Peña-Arellano, F.E.; Speake, C.C. Mirror tilt immunity interferometry with a cat's eye retroreflector. *Appl. Opt.* **2011**, *50*, 981–991.

Ryger, Ivan; Artusio-Glimpse, Alexandra; Williams, Paul; Shaw, Gordon; Simons, Matthew; Holloway, Christopher; Lehman, John.
"MEMS non-absorbing electromagnetic power sensor employing the effect of radiation pressure."
Paper presented at Eurosensors 2018, Graz, Austria. September 9, 2018 - September 12, 2018.

Pseudorandom Quantum States

Zhengfeng Ji[1], Yi-Kai Liu[2], and Fang Song[3]

[1] Centre for Quantum Software and Information, School of Software,
Faculty of Engineering and Information Technology,
University of Technology Sydney, NSW, Australia
Zhengfeng.Ji@uts.edu.au
[2] Joint Center for Quantum Information and Computer Science (QuICS),
National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA,
and University of Maryland, College Park, MD, USA
yi-kai.liu@nist.gov
[3] Computer Science Department, Portland State University, Portland, OR, USA
fang.song@pdx.edu

Abstract. We propose the concept of pseudorandom quantum states, which appear random to any quantum polynomial-time adversary. It offers a *computational* approximation to perfectly random quantum states analogous in spirit to cryptographic pseudorandom generators, as opposed to *statistical* notions of quantum pseudorandomness that have been studied previously, such as quantum t-designs analogous to t-wise independent distributions.

Under the assumption that quantum-secure one-way functions exist, we present efficient constructions of pseudorandom states, showing that our definition is achievable. We then prove several basic properties of pseudorandom states, which show the utility of our definition. First, we show a cryptographic no-cloning theorem: no efficient quantum algorithm can create additional copies of a pseudorandom state, when given polynomially-many copies as input. Second, as expected for random quantum states, we show that pseudorandom quantum states are highly entangled on average. Finally, as a main application, we prove that any family of pseudorandom states naturally gives rise to a private-key quantum money scheme.

1 Introduction

Pseudorandomness is a foundational concept in modern cryptography and theoretical computer science. A distribution $\mathcal{D}$, e.g., over a set of strings or functions, is called *pseudorandom* if no computationally-efficient observer can distinguish between an object sampled from $\mathcal{D}$, and a truly random object sampled from the uniform distribution [57,64,10]. Pseudorandom objects, such as pseudorandom generators (PRGs), pseudorandom functions (PRFs) and pseudorandom permutations (PRPs) are fundamental cryptographic building blocks, such as in the design of stream ciphers, block ciphers and message authentication codes [25,38,24,54,28]. Pseudorandomness is also essential in algorithm design and complexity theory such as derandomization [48,33].

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

The law of quantum physics asserts that truly random bits can be generated easily even with untrusted quantum devices [15,42]. Is pseudorandomness, a seemingly weaker notion of randomness, still relevant in the context of quantum information processing? The answer is yes. By a simple counting argument, one needs exponentially many bits even to specify a truly random function on n-bit strings. Hence, in the *computational* realm, pseudorandom objects that offer efficiency as well as other unique characteristics and strengths are indispensable.

A fruitful line of work on pseudorandomness in the context of quantum information science has been about quantum t-designs and unitary t-designs [4,17,27,12,16,34,60,70,46,45,44,11,41]. However, while these objects are often called "pseudorandom" in the mathematical physics literature, they are actually analogous to t-wise independent random variables in theoretical computer science. Our focus in this work is a notion of *computational* pseudorandomness, and in particular suits (complexity-theoretical) cryptography.

The major difference between t-wise independence and cryptographic pseudorandomness is the following. In the case of t-wise independence, the observer who receives the random-looking object may be computationally unbounded, but only *a priori* (when the random-looking object is constructed) fixed number t samples are given. Thus, quantum t-designs satisfy an "information-theoretic" or "statistical" notion of security. In contrast, in the case of cryptographic pseudorandomness, the observer who receives the random-looking object is assumed to be computationally efficient, in that it runs in probabilistic polynomial time for an arbitrary polynomial that is chosen by the observer, *after* the random-looking object has been constructed. This leads to a "computational" notion of security, which typically relies on some complexity-theoretic assumption, such as the existence of one-way functions).

In general, these two notions, t-wise independence and cryptographic pseudorandomness, are incomparable. In some ways, the setting of cryptographic pseudorandomness imposes stronger restrictions on the observer, since it assumes a bound on the observer's total computational effort (say, running in probabilistic polynomial time). In other ways, the setting of t-wise independence imposes stronger restrictions on the observer, since it forces the observer to make a limited number of non-adaptive "queries," specified by the parameter t, which is usually a constant or a fixed polynomial. In addition, different distance measures are often used, e.g., trace distance or diamond norm, versus computational distinguishability.

Cryptographic pseudorandomness in quantum information, which has received relatively less study, mostly connects with quantum money and post-quantum cryptography. Pseudorandomness is used more-or-less implicitly in quantum money, to construct quantum states that look complicated to a dishonest party, but have some hidden structure that allows them to be verified by the bank [1,40,2,3,69]. In post-quantum cryptography, one natural question is whether the classical constructions such as PRFs and PRPs remain secure against quantum attacks. This is a challenging task as, for example, a quantum adversary may query the underlying function or permutation in *superposition*. Fortunately, people have

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

so far restored several positive results. Assuming a one-way function that is hard to invert for polynomial-time quantum algorithms, we can attain quantum-secure PRGs as well as PRFs [28,66]. Furthermore, one can construct quantum-secure PRPs from quantum-secure PRFs using various *shuffling* constructions [68,58].

In this work, we study pseudorandom *quantum* objects such as quantum states and unitary operators. Quantum states (in analogy to strings) and unitary operations (in analogy to functions) form continuous spaces, and the Haar measure is considered the perfect randomness on the spaces of quantum states and unitary operators. A basic question is:

How to define and construct computational pseudorandom *approximations of Haar randomness, and what are their applications?*

Our contributions. We propose definitions of pseudorandom quantum states (PRS's) and pseudorandom unitary operators (PRUs), present efficient constructions of PRS's, demonstrate basic properties such as no-cloning and high entanglement of pseudorandom states, and showcase the construction of private-key quantum money schemes as one of the applications.

1. We propose a suitable definition of *quantum pseudorandom states*.
   We employ the notion of quantum *computational indistinguishability* to define quantum pseudorandom states. Loosely speaking, we consider a collection of quantum states $\left\{|\phi_k\rangle\right\}$ indexed by $k \in \mathcal{K}$, and require that no efficient quantum algorithm can distinguish between $|\phi_k\rangle$ for a random k and a state drawn according to the Haar measure. However, as a unique consideration in the quantum setting, we need to be cautious about *how many copies* of the input state are available to an adversary.
   Classically, this is a vacuous concern for defining a pseudorandom distribution on strings, since one can freely produce many copies of the input string. The quantum no-cloning theorem, however, forbids copying an unknown quantum state in general. Pseudorandom states in terms of *single-copy* indistinguishability have been discussed in the literature (see for example [13] and a recent study [14]). Though this single-copy definition may be suitable for certain cryptographic applications, it also loses many properties of Haar random states as a purely classical distributions already satisfies the definition [4].
   Therefore we require that no adversary can tell a difference even given any *polynomially many* copies of the state. This subsumes the single-copy version and is strictly stronger. We gain from it many interesting properties, such as the no-cloning property and entanglement property for pseudorandom states as discussed later in the paper.
2. We present concrete efficient constructions of PRS's with the minimal assumption that quantum-secure one-way functions exist.

[4] For example, a uniform distribution over the computational basis state $\{|k\rangle\}$ has an identical density matrix as a Haar random state and satisfy the single-shot definition of PRS. But distinguishing them becomes easy as soon as we have more than one copies. These states also do not appear to be hard to clone or possess entanglement.

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

Our construction uses any quantum-secure $\mathsf{PRF} = \{\mathsf{PRF}_k\}_{k\in\mathcal{K}}$ and computes it into the phases of a uniform superposition state (see equation (8)). We call such family of PRS the *random phase states*. This family of states can be efficiently generated using the quantum Fourier transform and a phase kick-back trick. We prove that this family of state is pseudorandom by a hybrid argument. By the quantum security of PRF, the family is computationally indistinguishable from a similar state family defined by truly random functions. We then prove that, this state family corresponding to truly random functions is statistically indistinguishable from Haar random states. Finally, by the fact that PRF exists assuming quantum-secure one-way functions, we can base our PRS construction on quantum-secure one-way functions.

We note that Aaronson [1, Theorem 3] has described a similar family of states, which uses some polynomial function instead of a PRF in the phases. In that construction, however, the size of the state family depends on (i.e., has to grow with) the adversary's number of queries that the family wants to tolerate. It therefore fails to satisfy our definition, in which any polynomial number queries independent of the family are permitted.

3. We prove *cryptographic no-cloning theorems* for PRS's, and they give a simple and generic construction of private-key quantum money schemes based on any PRS.

   We prove that a PRS remains pseudorandom, even if we additionally give the distinguisher an oracle that reflects about the given state (i.e., $O_\phi := \mathbb{1} - 2|\phi\rangle\langle\phi|$). This establishes the equivalence between the standard and a strong definition of PRS's. Technically, this is proved using the fact that with polynomially many copies of the state, one can approximately simulate the reflection oracle O_ϕ.

   We obtain general *cryptographic no-cloning theorems* of PRS's both with and without the reflection oracle. The theorems roughly state that given any polynomially many copies of pseudorandom states, no polynomial-time quantum algorithm can produce even one more copy of the state. We call them cryptographic no-cloning theorems due to the computational nature of our PRS. The proofs of these theorems use SWAP tests in the reduction from a hypothetical cloning algorithm to an efficient distinguishing algorithm violating the definition of PRS's.

   Using the strong pseudorandomness and the cryptographic no-cloning theorem with reflection oracle, we show that any PRS immediately gives a *private-key quantum money scheme*. While much attention has been focused on public-key quantum money [1,40,2,3,69], we emphasize that private-key quantum money is already non-trivial. Early schemes for private-key quantum money due to Wiesner and others were not *query secure*, and could be broken by online attacks [62,9,39,20]. Aaronson and Christiano finally showed a query-secure scheme in 2012, which achieves information-theoretic security in the random oracle model, and computational security in the standard model [2]. They used a specific construction based on hidden subspace states, whereas our construction (which is also query-secure) is more generic and can be based on any PRS. The freedom to choose and tweak the underlying

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

pseudorandom functions or permutations in the PRS may motivate and facilitate the construction of public-key quantum money schemes in future work.

4. We show that pseudorandom states are highly entangled.
   It is known that a Haar random state is entangled with high probability. We establish a similar result for any family of pseudorandom states. Namely, the states in any PRS family are entangled on average. It is shown that the expected Schmidt rank for any PRS is superpolynomial in κ and that the expected min entropy and von Neumann entropy are of the order $\omega(\log \kappa)$ where κ is the security parameter. This is yet another evidence of the suitability of our definition.
   The proof again rests critically on that our definition grants multiple copies to the distinguisher—if the expected entanglement is low, then SWAP test with respect to the corresponding subsystems of two copies of the state will serve as a distinguisher that violates the definition.
5. We propose a definition of *quantum pseudorandom unitary operators* (PRUs). We also present candidate constructions of PRUs (without a proof of security), by extending our techniques for constructing PRS's.
   Loosely speaking, these candidate PRUs resemble unitary t-designs that are constructed by interleaving random permutations with the quantum Fourier transform [27], or by interleaving random diagonal unitaries with the Hadamard transform [45,44], and iterating this construction several times. We conjecture that a PRU can be obtained in this way, using only a constant number of iterations. This is in contrast to unitary t-designs, where a parameter counting argument suggests that the number of iterations must grow with t. This conjecture is motivated by examples such as the Luby-Rackoff construction of a pseudorandom permutation using multi-round Feistel network built using a PRF.

Table 1. Summary of various notions that approximate true randomness

	Classical	**Quantum**
True randomness	Uniform distribution	Haar measure
t-wise independence	t-wise independent random variables	Quantum t-designs
Pseudorandomness	PRGs PRFs, PRPs	(*this work*) PRS's PRUs

Discussion. We summarize the mentioned variants of randomness in Table 1. The focus of this work is mostly about PRS's and we briefly touch upon PRUs. We view our work as an initial step and anticipate further fundamental investigation inspired by our notion of pseudorandom states and unitary operators.

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

We mention some immediate open problems. First, can we prove the security of our candidate PRU constructions? The techniques developed in quantum unitary designs [27,12] seem helpful. Second, are quantum-secure one-way functions necessary for the construction of PRS's? Third, can we establish security proofs for more candidate constructions of PRS's? Different constructions may have their own special properties that may be useful in different settings. It is also interesting to explore whether our quantum money construction may be adapted to a public-key money scheme under reasonable cryptographic assumptions. Finally, the entanglement property we prove here refers to the standard definitions of entanglement. If we approach the concept of pseudo-entanglement as a quantum analogue of pseudo-entropy for a distribution [7], can we improve the quantitative bounds?

We point out a possible application in physics. PRS's may be used in place of high-order quantum t-designs, giving a performance improvement in certain applications. For example, pseudorandom states can be used to construct toy models of quantum *thermalization*, where one is interested in quantum states that can be prepared efficiently via some dynamical process, yet have "generic" or "typical" properties as exemplified by Haar-random pure states, for instance [52]. Using t-designs with polynomially large t, one can construct states that are "generic" in a information-theoretic sense [36]. Using PRS, one can construct states that satisfy a weaker property: they are computationally indistinguishable from "generic" states, for a polynomial-time observer.

In these applications, PRS states may be more physically plausible than high-order quantum t-designs, because PRS states can be prepared in a shorter time, e.g., using a polylogarithmic-depth quantum circuit, based on known constructions for low-depth PRFs [6,47].

2 Preliminaries

2.1 Notions

For a finite set $\mathcal{X}$, $|\mathcal{X}|$ denotes the number of elements in $\mathcal{X}$. We use the notion $\mathcal{Y}^{\mathcal{X}}$ to denote the set of all functions $f : \mathcal{X} \to \mathcal{Y}$. For finite set $\mathcal{X}$, we use $x \leftarrow \mathcal{X}$ to mean that x is drawn uniformly at random from $\mathcal{X}$. The permutation group over elements in $\mathcal{X}$ is denoted as $S_{\mathcal{X}}$. We use $\mathrm{poly}(\kappa)$ to denote the collection of polynomially bounded functions of the security parameter κ, and use $\mathrm{negl}(\kappa)$ to denote negligible functions in κ. A function $\epsilon(\kappa)$ is *negligible* if for all constant $c > 0$, $\epsilon(\kappa) < \kappa^{-c}$ for large enough κ.

In this paper, we use a *quantum register* to name a collection of qubits that we view as a single unit. Register names are represented by capital letters in a *sans serif* font. We use $\mathrm{S}(\mathcal{H})$, $\mathrm{D}(\mathcal{H})$, $\mathrm{U}(\mathcal{H})$ and $\mathrm{L}(\mathcal{H})$ to denote the set of pure quantum states, density operators, unitary operators and bounded linear operators on space $\mathcal{H}$ respectively. An ensemble of states $\{(p_i, \rho_i)\}$ represents a system prepared in ρ_i with probability p_i. If the distribution is uniform, we write the ensemble as $\{\rho_i\}$. The adjoint of matrix M is denoted as M^*. For matrix

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

M, $|M|$ is defined to be $\sqrt{M^*M}$. The operator norm $\|M\|$ of matrix M is the largest eigenvalue of $|M|$. The trace norm $\|M\|_1$ of M is the trace of $|M|$. For two operators $M, N \in \mathrm{L}(\mathcal{H})$, the Hilbert-Schmidt inner product is defined as

$$\langle M, N\rangle = \mathrm{tr}(M^*N).$$

A quantum channel is a physically admissible transformation of quantum states. Mathematically, a quantum channel

$$\mathcal{E} : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{K})$$

is a completely positive, trace-preserving linear map.

The trace distance of two quantum states $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{H})$ is

$$\mathrm{TD}(\rho_0, \rho_1) \stackrel{\mathrm{def}}{=} \frac{1}{2}\left\|\rho_0 - \rho_1\right\|_1. \tag{1}$$

It is known (Holevo-Helstrom theorem [30,31]) that for a state drawn uniformly at random from the set $\{\rho_0, \rho_1\}$, the optimal distinguish probability is given by

$$\frac{1 + \mathrm{TD}(\rho_0, \rho_1)}{2}.$$

Define number $N = 2^n$ and set $\mathcal{X} = \{0, 1, \ldots, N-1\}$. The quantum Fourier transform on n qubits is defined as

$$F = \frac{1}{\sqrt{N}} \sum_{x,y\in\mathcal{X}} \omega_N^{xy} |x\rangle\langle y|. \tag{2}$$

It is a well-known fact in quantum computing that F can be implemented in time $\mathrm{poly}(n)$.

For Hilbert space $\mathcal{H}$ and integer m, we use $\vee^m\mathcal{H}$ to denote the symmetric subspace of $\mathcal{H}^{\otimes m}$, the subspace of states that are invariant under permutations of the subsystems. Let N be the dimension of $\mathcal{H}$ and let $\mathcal{X}$ be the set $\{0, 1, \ldots, N-1\}$ such that $\mathcal{H}$ is the span of $\{|x\rangle\}_{x\in\mathcal{X}}$. For any $\mathbf{x} = (x_1, x_2, \ldots, x_m) \in \mathcal{X}^m$, let m_j be the number of j in $\mathbf{x}$ for $j \in \mathcal{X}$. Define state

$$\big|\mathbf{x}; \mathrm{Sym}\big\rangle = \sqrt{\frac{\prod_{j\in\mathcal{X}} m_j!}{m!}} \sum_{\sigma} \Big| x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m)} \Big\rangle. \tag{3}$$

The summation runs over all possible permutations σ that give different tuples $(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m)})$. Equivalently, we have

$$\big|\mathbf{x}; \mathrm{Sym}\big\rangle = \frac{1}{\sqrt{m! \prod_{j\in\mathcal{X}} m_j!}} \sum_{\sigma\in S_m} \Big| x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m)} \Big\rangle. \tag{4}$$

The coefficients in the front of the above two equations are normalization constants. The set of states

$$\big\{\big|\mathbf{x}; \mathrm{Sym}\big\rangle\big\}_{\mathbf{x}\in\mathcal{X}^m} \tag{5}$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

forms an orthonormal basis of the symmetric subspace $\vee^m \mathcal{H}$ [59, Prop.7.2]. This implies that the dimension of the symmetric subspace is

$$\binom{N+m-1}{m}.$$

Let Π_m^{Sym} be the projection onto the symmetric subspace $\vee^m \mathcal{H}$. For a permutation $\sigma \in S_m$, define operator

$$W_\sigma = \sum_{x_1,x_2,\ldots,x_m \in \mathcal{X}} \left|x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, \ldots, x_{\sigma^{-1}(m)}\right\rangle\!\left\langle x_1, x_2, \ldots, x_m\right|.$$

The following identity will be useful [59, Prop.7.1]

$$\Pi_m^{\mathrm{Sym}} = \frac{1}{m!} \sum_{\sigma \in S_m} W_\sigma. \tag{6}$$

Let μ be the Haar measure on $\mathrm{S}(\mathcal{H})$, it is known that [26, Prop.6]

$$\int \left(|\psi\rangle\langle\psi|\right)^{\otimes m} \mathrm{d}\mu(\psi) = \binom{N+m-1}{m}^{-1} \Pi_m^{\mathrm{Sym}}. \tag{7}$$

2.2 Cryptography

In this section, we recall several definitions and results from cryptography that is necessary for this work.

Pseudorandom functions (PRF) and pseudorandom permutations (PRP) are important constructions in classical cryptography. Intuitively, they are families of functions or permutations that looks like truly random functions or permutations to polynomial-time machines. In the quantum case, we need a strong requirement that they still look random even to polynomial-time quantum algorithms.

Definition 1 (Quantum-Secure Pseudorandom Functions and Permutations). *Let $\mathcal{K}$, $\mathcal{X}$, $\mathcal{Y}$ be the key space, the domain and range, all implicitly depending on the security parameter κ. A keyed family of functions $\left\{\mathsf{PRF}_k : \mathcal{X} \to \mathcal{Y}\right\}_{k\in\mathcal{K}}$ is a quantum-secure pseudorandom function (QPRF) if for any polynomial-time quantum oracle algorithm $\mathcal{A}$, PRF_k with a random $k \leftarrow \mathcal{K}$ is indistinguishable from a truly random function $f \leftarrow \mathcal{Y}^{\mathcal{X}}$ in the sense that:*

$$\left|\Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}^{\mathsf{PRF}_k}(1^\kappa) = 1\right] - \Pr_{f\leftarrow\mathcal{Y}^{\mathcal{X}}}\left[\mathcal{A}^{f}(1^\kappa) = 1\right]\right| = \mathrm{negl}(\kappa).$$

Similarly, a keyed family of permutations $\left\{\mathsf{PRP}_k \in S_{\mathcal{X}}\right\}_{k\in\mathcal{K}}$ is a quantum-secure pseudorandom permutation (QPRP) if for any quantum algorithm $\mathcal{A}$ making at most polynomially many queries, PRP_k with a random $k \leftarrow \mathcal{K}$ is indistinguishable from a truly random permutation in the sense that:

$$\left|\Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}^{\mathsf{PRP}_k}(1^\kappa) = 1\right] - \Pr_{P\leftarrow S_{\mathcal{X}}}\left[\mathcal{A}^{P}(1^\kappa) = 1\right]\right| = \mathrm{negl}(\kappa).$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

In addition, both PRF_k and PRP_k are polynomial-time computable (on a classical computer).

Fact 1 *QPRFs and QPRPs exist if quantum-secure one-way functions exist.*

Zhandry proved the existence of QPRFs assuming the existence of one-way functions that are hard to invert even for quantum algorithms [66]. Assuming QPRF, one can construct QPRP using various *shuffling* constructions [68,58]. Since a random permutation and a random function is indistinguishable by efficient quantum algorithms [65,67], existence of QPRP is hence equivalent to existence of QPRF.

3 Pseudorandom Quantum States

In this section, we will discuss the definition and constructions of pseudorandom quantum states.

3.1 Definition of Pseudorandom States

Intuitively speaking, a family pseudorandom quantum states are a set of random states $\left\{|\phi_k\rangle\right\}_{k\in\mathcal{K}}$ that is indistinguishable from Haar random quantum states.

The first idea on defining pseudorandom states can be the following. Without loss of generality, we consider states in $\mathrm{S}(\mathcal{H})$ where $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ is the Hilbert space for n-qubit systems. We are given either a state randomly sampled from the set $\left\{|\phi_k\rangle \in \mathcal{H}\right\}_{k\in\mathcal{K}}$ or a state sampled according to the Haar measure on $\mathrm{S}(\mathcal{H})$, and we require that no efficient quantum algorithm will be able to tell the difference between the two cases.

However, this definition does not seem to grasp the quantum nature of the problem. First, the state family where each $|\phi_k\rangle$ is a uniform random bit string will satisfy the definition—in both cases, the mixed states representing the ensemble are $\mathbb{1}/2^n$. Second, many of the applications that we can find for PRS's will not hold for this definition.

Instead, we require that the family of states looks random even if polynomially many copies of the state are given to the distinguishing algorithm. We argue that this is the more natural way to define pseudorandom states. One can see that this definition also naturally generalizes the definition of pseudorandomness in the classical case to the quantum setting. In the classical case, asking for more copies of a string is always possible and one does not bother making this explicit in the definition. This of course also rules out the example of classical random bit strings we discussed before. Moreover, this strong definition, once established, is rather flexible to use when studying the properties and applications of pseudorandom states.

Definition 2 (Pseudorandom Quantum States (PRS's)). *Let κ be the security parameter. Let $\mathcal{H}$ be a Hilbert space and $\mathcal{K}$ the key space, both parameterized by κ. A keyed family of quantum states $\left\{|\phi_k\rangle \in \mathrm{S}(\mathcal{H})\right\}_{k\in\mathcal{K}}$ is* ***pseudorandom****, if the following two conditions hold:*

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

1. *(**Efficient generation**). There is a polynomial-time quantum algorithm G that generates state $|\phi_k\rangle$ on input k. That is, for all $k \in \mathcal{K}$, $G(k) = |\phi_k\rangle$.*
2. *(**Pseudorandomness**). Any polynomially many copies of $|\phi_k\rangle$ with the same random $k \in \mathcal{K}$ is **computationally indistinguishable** from the same number of copies of a Haar random state. More precisely, for any efficient quantum algorithm $\mathcal{A}$ and any $m \in \mathrm{poly}(\kappa)$,*

$$\left| \Pr_{k \leftarrow \mathcal{K}} \left[\mathcal{A}(|\phi_k\rangle^{\otimes m}) = 1 \right] - \Pr_{|\psi\rangle \leftarrow \mu} \left[\mathcal{A}(|\psi\rangle^{\otimes m}) = 1 \right] \right| = \mathrm{negl}(\kappa),$$

where μ is the Haar measure on $\mathrm{S}(\mathcal{H})$.

3.2 Constructions and Analysis

In this section, we give an efficient construction of pseudorandom states which we call random phase states. We will prove that this family of states satisfies our definition of PRS's. There are other interesting and simpler candidate constructions, but the family of random phase states is the easiest to analyze.

Let $\mathsf{PRF} : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$ be a quantum-secure pseudorandom function with key space $\mathcal{K}$, $\mathcal{X} = \{0, 1, 2, \ldots, N-1\}$ and $N = 2^n$. $\mathcal{K}$ and N are implicitly functions of the security parameter κ. The family of pseudorandom states of n qubits is defined

$$|\phi_k\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \mathcal{X}} \omega_N^{\mathsf{PRF}_k(x)} |x\rangle, \tag{8}$$

for $k \in \mathcal{K}$ and $\omega_N = \exp(2\pi i/N)$.

Theorem 1. *For any QPRF $\mathsf{PRF} : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$, the family of states $\{|\phi_k\rangle\}_{k \in \mathcal{K}}$ defined in Eq. (8) is a PRS.*

Proof. First, we prove that the state can be efficiently prepared with a single query to PRF_k. As PRF_k is efficient, this proves the efficient generation property.

The state generation algorithm works as follows. First, it prepares a state

$$\frac{1}{N} \sum_{x \in \mathcal{X}} |x\rangle \sum_{y \in \mathcal{X}} \omega_N^y |y\rangle.$$

This can be done by applying $H^{\otimes n}$ to the first register initialized in $|0\rangle$ and the quantum Fourier transform to the second register in state $|1\rangle$.

Then the algorithm calls PRF_k on the first register and subtract the result from the second register, giving state

$$\frac{1}{N} \sum_{x \in \mathcal{X}} |x\rangle \sum_{y \in \mathcal{X}} \omega_N^y |y - \mathsf{PRF}_k(x)\rangle.$$

The state can be rewritten as

$$\frac{1}{N} \sum_{x \in \mathcal{X}} \omega_N^{\mathsf{PRF}_k(x)} |x\rangle \sum_{y \in \mathcal{X}} \omega_N^y |y\rangle.$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

Therefore, the effect of this step is to transform the first register to the required form and leaving the second register intact.

Next, we prove the pseudorandomness property of the family. For this purpose, we consider three hybrids. In the first hybrid H_1, the state will be $|\phi_k\rangle^{\otimes m}$ for a uniform random $k \in \mathcal{K}$. In the second hybrid H_2, the state is $|f\rangle^{\otimes m}$ for truly random functions $f \in \mathcal{X}^{\mathcal{X}}$ where

$$|f\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \mathcal{X}} \omega_N^{f(x)} |x\rangle.$$

In the third hybrid H_3, the state is $|\psi\rangle^{\otimes m}$ for $|\psi\rangle$ chosen according to the Haar measure.

By the definition of the quantum-secure pseudorandom functions for PRF, we have for any polynomial-time quantum algorithm $\mathcal{A}$ and any $m \in \mathrm{poly}(\kappa)$,

$$\big|\Pr\big[\mathcal{A}(H_1) = 1\big] - \Pr\big[\mathcal{A}(H_2) = 1\big]\big| = \mathrm{negl}(\kappa).$$

By Lemma 1, we have for any algorithm $\mathcal{A}$ and $m \in \mathrm{poly}(\kappa)$,

$$\big|\Pr\big[\mathcal{A}(H_2) = 1\big] - \Pr\big[\mathcal{A}(H_3) = 1\big]\big| = \mathrm{negl}(\kappa).$$

This completes the proof by triangle inequality.

Lemma 1. *For function $f : \mathcal{X} \to \mathcal{X}$, define quantum state*

$$|f\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \mathcal{X}} \omega_N^{f(x)} |x\rangle.$$

For $m \in \mathrm{poly}(\kappa)$, the state ensemble $\big\{|f\rangle^{\otimes m}\big\}$ is statistically indistinguishable from $\big\{|\psi\rangle^{\otimes m}\big\}$ for Haar random $|\psi\rangle$.

Proof. Let $m \in \mathrm{poly}(\kappa)$ be the number of copies of the state. We have

$$\mathop{\mathbb{E}}_f\Big[\big(|f\rangle\langle f|\big)^{\otimes m}\Big] = \frac{1}{N^m} \sum_{\mathbf{x} \in \mathcal{X}^m, \mathbf{y} \in \mathcal{X}^m} \mathop{\mathbb{E}}_f \omega_N^{f(x_1)+\cdots+f(x_m)-[f(y_1)+\cdots+f(y_m)]} \big|\mathbf{x}\big\rangle\big\langle\mathbf{y}\big|,$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_m)$. For later convenience, define density matrix

$$\rho^m = \mathop{\mathbb{E}}_f\Big[\big(|f\rangle\langle f|\big)^{\otimes m}\Big].$$

We will compute the entries of ρ^m explicitly.

For $\mathbf{x} = (x_1, x_2, \ldots, x_m) \in \mathcal{X}^m$, let m_j be the number of j in $\mathbf{x}$ for $j \in \mathcal{X}$. Obviously, one has $\sum_{j \in \mathcal{X}} m_j = m$. Note that we have omitted the dependence of m_j on $\mathbf{x}$ for simplicity. Recall the basis states defined in Eq. (4)

$$\big|\mathbf{x}; \mathrm{Sym}\big\rangle = \frac{1}{\sqrt{\Big(\prod_{j \in \mathcal{X}} m_j!\Big) m!}} \sum_{\sigma \in S_m} \Big|x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m)}\Big\rangle.$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

For $\mathbf{x},\mathbf{y}\in\mathcal{X}^m$, let m_j be the number of j in $\mathbf{x}$ and m'_j be the number of j in $\mathbf{y}$.

We can compute the entries of ρ^m as

$$\begin{aligned}&\big\langle \mathbf{x};\mathrm{Sym}\big|\rho^m\big|\mathbf{y};\mathrm{Sym}\big\rangle\\ =&\frac{m!}{N^m\sqrt{\Big(\prod_{j\in\mathcal{X}} m_j!\Big)\Big(\prod_{j\in\mathcal{X}} m'_j!\Big)}}\mathop{\mathbb{E}}_f\Big[\exp\Big(\frac{2\pi i}{N}\sum_{l=1}^m\big(f(x_l)-f(y_l)\big)\Big)\Big].\end{aligned}$$

When $\mathbf{x}$ is not a permutation of $\mathbf{y}$, the summation $\sum_{l=1}^m\big(f(x_l)-f(y_l)\big)$ is a summation of terms $\pm f(z_j)$ for distinct values z_j. As f is a truly random function, $f(z_j)$ is uniformly random and independent of $f(z_{j'})$ for $z_j\neq z_{j'}$. So it is not hard to verify that the entry is nonzero only if $\mathbf{x}$ is a permutation of $\mathbf{y}$. These nonzero entries are on the diagonal of ρ^m in the basis of $\big\{\big|\mathbf{x};\mathrm{Sym}\big\rangle\big\}$. These diagonal entries are

$$\big\langle \mathbf{x};\mathrm{Sym}\big|\rho^m\big|\mathbf{x};\mathrm{Sym}\big\rangle=\frac{m!}{N^m\prod_{j\in\mathcal{X}} m_j!}\,.$$

Let ρ_μ^m be the density matrix of a random state $|\psi\rangle^{\otimes m}$, for $|\psi\rangle$ chosen from the Haar measure μ. From Eqs. (5) and (7), we have that

$$\rho_\mu^m=\binom{N+m-1}{m}^{-1}\sum_{\mathbf{x};\mathrm{Sym}}\big|\mathbf{x};\mathrm{Sym}\big\rangle\big\langle\mathbf{x};\mathrm{Sym}\big|.$$

We need to prove

$$\mathrm{TD}\big(\rho^m,\rho_\mu^m\big)=\mathrm{negl}(\kappa).$$

Define

$$\delta_{\mathbf{x};\mathrm{Sym}}=\frac{m!}{N^m\prod_{j\in\mathcal{X}} m_j!}-\binom{N+m-1}{m}^{-1}.$$

Then

$$\mathrm{TD}(\rho^m,\rho_\mu^m)=\frac{1}{2}\sum_{\mathbf{x};\mathrm{Sym}}\big|\delta_{\mathbf{x};\mathrm{Sym}}\big|\,.$$

The ratio of the two terms in $\delta_{\mathbf{x};\mathrm{Sym}}$ is

$$\frac{m!\binom{N+m-1}{m}}{N^m\prod\limits_{j\in\mathcal{X}} m_j!}=\frac{\prod\limits_{l=0}^{m-1}\Big(1+\frac{l}{N}\Big)}{\prod\limits_{j\in\mathcal{X}} m_j!}.$$

For sufficient large security parameter κ, the ratio is larger than 1 only if $\prod_{j\in\mathcal{X}} m_j!=1$, which corresponds to $\mathbf{x}$'s whose entries are all distinct. As there are $\binom{N}{m}$ such $\mathbf{x}$'s, we can calculate the trace distance as

$$\begin{aligned}\mathrm{TD}\big(\rho^m,\rho_\mu^m\big)&=\binom{N}{m}\left[\frac{m!}{N^m}-\binom{N+m-1}{m}^{-1}\right]\\&=\frac{N(N-1)\cdots(N-m+1)}{N^m}-\frac{N(N-1)\cdots(N-m+1)}{(N+m-1)\cdots N}.\end{aligned}$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

As first term is less than 1 and is at least

$$(1-\frac{1}{N})\cdots(1-\frac{m-1}{N}) \geq 1-\frac{1+2+\cdots+(m-1)}{N}$$

For our choices of $m \in \text{poly}(\kappa)$ and $N \in 2^{\text{poly}(\kappa)}$, this term is $1-\text{negl}(\kappa)$ for sufficiently large security parameter κ. Similar analysis applies to the second term and this completes the proof.

3.3 Comparison with Related Work

We remark that a similar family of states was considered in [1] (Theorem 3). However, the size of the state family there depends on a parameter d which should be larger than the sum of the number of state copies and the number of queries. In our construction, the key space is fixed for a given security parameter, which may be advantageous for various applications.

We mention several other candidate constructions of PRS's and leave detailed analysis of them to future work. A construction closely related to the random phase states in Eq. (8) uses random ± 1 phases,

$$|\phi_k\rangle = \frac{1}{\sqrt{N}}\sum_{x\in\mathcal{X}}(-1)^{\mathsf{PRF}_k(x)}|x\rangle.$$

Intuitively, this family is less random than the random phase states in Eq. (8) and the corresponding density matrix ρ^m has small off-diagonal entries, making the proof more challenging. The other family of candidate states on $2n$ qubits takes the form

$$|\phi_k\rangle = \frac{1}{\sqrt{N}}\mathsf{PRP}_k\left[\sum_{x\in\mathcal{X}}|x\rangle\otimes|0^n\rangle\right].$$

In this construction, the state is an equal superposition of a random subset of size 2^n of $\{0,1\}^{2n}$ and PRP is any pseudorandom permutation over the set $\{0,1\}^{2n}$. We call this the *random subset states* construction.

Finally, we remark that under plausible cryptographic assumptions our PRS constructions can be implemented using shallow quantum circuits of polylogarithmic depth. To see this, note that there exist PRFs that can be computed in polylogarithmic depth [6], which are based on lattice problems such as "learning with errors" (LWE) [53], and are believed to be secure against quantum computers. These PRFs can be used directly in our PRS construction. (Alternatively, one can use low-depth PRFs that are constructed from more general assumptions, such as the existence of trapdoor one-way permutations [47].)

This shows that PRS states can be prepared in surprisingly small depth, compared to quantum state t-designs, which generally require at least linear depth when t is a constant greater than 2, or polynomial depth when t grows polynomially with the number of qubits [4,12,44,41]. (Note, however, that for $t = 2$, quantum state 2-designs can be generated in logarithmic depth [16].) Moreover, PRS states are sufficient for many applications where high-order t-designs are used [52,36], provided that one only requires states to be *computationally* (not statistically) indistinguishable from Haar-random.

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

4 Cryptographic No-cloning Theorem and Quantum Money

A fundamental fact in quantum information theory is that unknown or random quantum states cannot be cloned [63,18,61,49,51]. The main topic of this section is to investigate the cloning problem for pseudorandom states. As we will see, even though pseudorandom states can be efficiently generated, they do share the no-cloning property of generic quantum states.

Let $\mathcal{H}$ be the Hilbert space of dimension N and $m < m'$ be two integers. The numbers N, m, m' depend implicitly on a security parameter κ. We will assume that N is exponential in κ and $m \in \text{poly}(\kappa)$ in the following discussion.

We first recall the fact that for Haar random state $|\psi\rangle \in \mathrm{S}(\mathcal{H})$, the success probability of producing m' copies of the state given m copies is negligibly small. Let $\mathcal{C}$ be a cloning channel that on input $(|\psi\rangle\langle\psi|)^{\otimes m}$ tries to output a state that is close to $(|\psi\rangle\langle\psi|)^{\otimes m'}$ for $m' > m$. The expected success probability of $\mathcal{C}$ is measured by

$$\int \Big\langle \big(|\psi\rangle\langle\psi|\big)^{\otimes m'}, \mathcal{C}\big(\big(|\psi\rangle\langle\psi|\big)^{\otimes m}\big) \Big\rangle \, \mathrm{d}\mu(\psi).$$

It is known that [61], for all cloning channel $\mathcal{C}$, this success probability is bounded by

$$\binom{N+m-1}{m} \Big/ \binom{N+m'-1}{m'},$$

which is $\text{negl}(\kappa)$ for our choices of N, m, m'.

We establish a no-cloning theorem for PRS's which says that no efficient quantum cloning procedure exists for a general PRS. The theorem is called the cryptographic no-cloning theorem because of its deep roots in pseudorandomness in cryptography.

Theorem 2 (Cryptographic No-cloning Theorem). *For any PRS family $\{|\phi_k\rangle\}_{k\in\mathcal{K}}$, $m \in \text{poly}(\kappa)$, $m < m'$ and any polynomial-time quantum algorithm $\mathcal{C}$, the success cloning probability*

$$\mathop{\mathbb{E}}_{k\in\mathcal{K}} \Big\langle \big(|\phi_k\rangle\langle\phi_k|\big)^{\otimes m'}, \mathcal{C}\big(\big(|\phi_k\rangle\langle\phi_k|\big)^{\otimes m}\big) \Big\rangle = \text{negl}(\kappa).$$

Proof. Assume on the contrary that there is a polynomial-time quantum cloning algorithm $\mathcal{C}$ such that the success cloning probability of producing $m+1$ from m copies is κ^{-c} for some constant $c > 0$. We will construct a polynomial-time distinguisher $\mathcal{D}$ that violates the definition of PRS's. Distinguisher $\mathcal{D}$ will draw $2m+1$ copies of the state, call $\mathcal{C}$ on the first m copies, and perform the SWAP test on the output of $\mathcal{C}$ and the remaining $m+1$ copies. It is easy to see that $\mathcal{D}$ outputs 1 with probability $(1+\kappa^{-c})/2$ if the input is from PRS, while if the input is Haar random, it outputs 1 with probability $(1+\text{negl}(\kappa))/2$. Since $\mathcal{C}$ is polynomial-time, it follows that $\mathcal{D}$ is also polynomial-time. This is a contradiction with the definition of PRS's and completes the proof.

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

4.1 A Strong Notion of PRS and Equivalence to PRS

In this section, we show that, somewhat surprisingly, PRS in fact implies a seemingly stronger notion, where indistinguishability needs to hold even if a distinguisher additionally has access to an oracle that reflects about the given state. There are at least a couple of motivations to consider an augmented notion. Firstly, unlike a classical string, a quantum state is inherently *hidden*. Give a quantum register prepared in some state (i.e., a physical system), we can only choose some observable to measure which just reveals partial information and will collapse the state in general. Therefore, it is meaningful to consider offering a distinguishing algorithm more information *describing* the given state, and the reflection oracle comes naturally. Secondly, this stronger notion is extremely useful in our application of quantum money schemes, and could be interesting elsewhere too.

More formally, for any state $|\phi\rangle \in \mathcal{H}$, define an oracle $O_\phi := \mathbb{1} - 2|\phi\rangle\langle\phi|$ that reflects about $|\phi\rangle$.

Definition 3 (Strongly Pseudorandom Quantum States). *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{K}$ be the key space. $\mathcal{H}$ and $\mathcal{K}$ depend on the security parameter κ. A keyed family of quantum states $\left\{|\phi_k\rangle \in \mathrm{S}(\mathcal{H})\right\}_{k\in\mathcal{K}}$ is **strongly pseudorandom**, if the following two conditions hold:*

1. *(**Efficient generation**). There is a polynomial-time quantum algorithm G that generates state $|\phi_k\rangle$ on input k. That is, for all $k \in \mathcal{K}$, $G(k) = |\phi_k\rangle$.*
2. *(**Strong Pseudorandomness**). Any polynomially many copies of $|\phi_k\rangle$ with the same random $k \in \mathcal{K}$ is **computationally indistinguishable** from the same number of copies of a Haar random state. More precisely, for any efficient quantum oracle algorithm $\mathcal{A}$ and any $m \in \mathrm{poly}(\kappa)$,*

$$\left|\Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}^{O_{\phi_k}}(|\phi_k\rangle^{\otimes m}) = 1\right] - \Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{A}^{O_\psi}(|\psi\rangle^{\otimes m}) = 1\right]\right| = \mathrm{negl}(\kappa),$$

where μ is the Haar measure on $\mathrm{S}(\mathcal{H})$.

Note that since the distinguisher $\mathcal{A}$ is polynomial-time, the number of queries to the reflection oracle (O_{ϕ_k} or O_ψ) is also polynomially bounded.

We prove the advantage that a reflection oracle may give to a distinguisher is limited. In fact, standard PRS implies strong PRS, and hence they are equivalent.

Theorem 3. *A family of states $\left\{|\phi_k\rangle\right\}_{k\in\mathcal{K}}$ is strongly pseudorandom if and only if it is (standard) pseudorandom.*

Proof. Clearly a strong PRS is also a standard PRS by definition. It suffice to prove that any PRS is also strongly pseudorandom.

Suppose for contradiction that there is a distinguishing algorithm $\mathcal{A}$ that breaks the strongly pseudorandom condition. Namely, there exists $m \in \mathrm{poly}(\kappa)$ and constant $c > 0$ such that for sufficiently large κ,

$$\left|\Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}^{O_{\phi_k}}(|\phi_k\rangle^{\otimes m}) = 1\right] - \Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{A}^{O_\psi}(|\psi\rangle^{\otimes m}) = 1\right]\right| = \varepsilon(\kappa) \geq \kappa^{-c}.$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

We assume $\mathcal{A}$ makes $q \in \mathrm{poly}(\kappa)$ queries to the reflection oracle. Then, by Theorem 4, there is an algorithm $\mathcal{B}$ such that for any l

$$\left| \Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}^{O_{\phi_k}}(|\phi_k\rangle^{\otimes m})\right] - \Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{B}(|\phi_k\rangle^{\otimes(m+l)})\right] \right| \leq \frac{2q}{\sqrt{l+1}},$$

and

$$\left| \Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{A}^{O_{\psi}}(|\psi\rangle^{\otimes m})\right] - \Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{B}(|\psi\rangle^{\otimes(m+l)})\right] \right| \leq \frac{2q}{\sqrt{l+1}}.$$

By triangle inequality, we have

$$\left| \Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{B}(|\phi_k\rangle^{\otimes(m+l)})\right] - \Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{B}(|\psi\rangle^{\otimes(m+l)})\right] \right| \geq \kappa^{-c} - \frac{4q}{\sqrt{l+1}}.$$

Choosing $l = 64q^2\kappa^{2c} \in \mathrm{poly}(\kappa)$, we have

$$\left| \Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{B}(|\phi_k\rangle^{\otimes(m+l)})\right] - \Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{B}(|\psi\rangle^{\otimes(m+l)})\right] \right| \geq \kappa^{-c}/2,$$

which is a contradiction with the definition of PRS for $\{|\phi_k\rangle\}$. Therefore, we conclude that PRS and strong PRS are equivalent.

We now show a technical ingredient that allows us to simulate the reflection oracle about a state by using multiple copies of the given state. This result is inspired by a similar theorem proved by Ambainis et al. [5, Lemma 42]. Our simulation applies the reflection about the standard symmetric subspace, as opposed to a reflection operation about a particular subspace in [5], on the multiple copies of the given state, which we know how to implement efficiently.

Theorem 4. *Let $|\psi\rangle \in \mathcal{H}$ be a quantum state. Define oracle $O_\psi = \mathbb{1} - 2|\psi\rangle\langle\psi|$ to be the reflection about $|\psi\rangle$. Let $|\xi\rangle$ be a state not necessarily independent of $|\psi\rangle$. Let $\mathcal{A}^{O_\psi}$ be an oracle algorithm that makes q queries to O_ψ. For any integer $l > 0$, there is a quantum algorithm $\mathcal{B}$ that makes no queries to O_ψ such that*

$$\mathrm{TD}\big(\mathcal{A}^{O_\psi}(|\xi\rangle), \mathcal{B}(|\psi\rangle^{\otimes l} \otimes |\xi\rangle)\big) \leq \frac{q\sqrt{2}}{\sqrt{l+1}}\,.$$

Moreover, the running time of $\mathcal{B}$ is polynomial in that of $\mathcal{A}$ and l.

Proof. Consider a quantum register T, initialized in the state $|\Theta\rangle_{\mathsf{T}} = |\psi\rangle^{\otimes l} \in \mathcal{H}^{\otimes l}$. Let Π be the projection onto the symmetric subspace $\vee^{l+1}\mathcal{H} \subset \mathcal{H}^{\otimes(l+1)}$, and let $R = \mathbb{1} - 2\Pi$ be the reflection about the symmetric subspace.

Assume without loss of generality that algorithm $\mathcal{A}$ is unitary and only performs measurements at the end. We define algorithm $\mathcal{B}$ to be the same as $\mathcal{A}$, except that when $\mathcal{A}$ queries O_ψ on register D, $\mathcal{B}$ applies the reflection R on the collection of quantum registers D and T. We first analyze the corresponding states after the first oracle call to O_ψ in algorithms $\mathcal{A}$ and $\mathcal{B}$,

$$|\Psi_A\rangle = O_\psi\big(|\phi\rangle_{\mathsf{D}}\big) \otimes |\Theta\rangle_{\mathsf{T}}, \quad |\Psi_B\rangle = R\big(|\phi\rangle_{\mathsf{D}} \otimes |\Theta\rangle_{\mathsf{T}}\big).$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

For any two states $|x\rangle, |y\rangle \in \mathcal{H}$, we have

$$\begin{aligned}\big(\langle x| \otimes \langle \Theta|\big) R \big(|y\rangle \otimes |\Theta\rangle\big) &= \langle x|y\rangle - 2 \mathop{\mathbb{E}}_{\pi \in S_{l+1}} \big(\langle x| \otimes \langle \Theta|\big) W_\pi \big(|y\rangle \otimes |\Theta\rangle\big) \\ &= \langle x|y\rangle - \frac{2}{l+1} \langle x|y\rangle - \frac{2l}{l+1} \langle x|\psi\rangle \langle \psi|y\rangle \\ &= \frac{l-1}{l+1} \langle x|y\rangle - \frac{2l}{l+1} \langle x|\psi\rangle \langle \psi|y\rangle ,\end{aligned}$$

where the first step uses the identity in Eq. (6) and the second step follows by observing that the probability of a random $\pi \in S_{l+1}$ mapping 1 to 1 is $1/(l+1)$. These calculations imply that,

$$\big(\mathbb{1} \otimes \langle \Theta|\big) R \big(\mathbb{1} \otimes |\Theta\rangle\big) = \frac{l-1}{l+1} \mathbb{1} - \frac{2l}{l+1} |\psi\rangle\langle\psi|.$$

We can compute the inner product of the two states $|\Psi_A\rangle$ and $|\Psi_B\rangle$ as

$$\begin{aligned}\langle \Psi_A|\Psi_B\rangle &= \mathrm{tr}\Big(\big(|\phi\rangle \otimes |\Theta\rangle\big) \big(\langle \phi| \otimes \langle \Theta|\big) \left(O_\psi \otimes \mathbb{1}\right) R \Big) \\ &= \mathrm{tr}\Big(|\phi\rangle\langle\phi| \, O_\psi \left(\mathbb{1} \otimes \langle \Theta|\right) R \left(\mathbb{1} \otimes |\Theta\rangle\right) \Big) \\ &= \mathrm{tr}\left(|\phi\rangle\langle\phi| \big(\mathbb{1} - 2|\psi\rangle\langle\psi|\big) \left(\frac{l-1}{l+1} \mathbb{1} - \frac{2l}{l+1} |\psi\rangle\langle\psi| \right) \right) \\ &= \frac{l-1}{l+1} + \frac{2l}{l+1} \left|\langle \phi|\psi\rangle\right|^2 - \frac{2(l-1)}{l+1} \left|\langle \phi|\psi\rangle\right|^2 \\ &= \frac{l-1}{l+1} + \frac{2}{l+1} \left|\langle \phi|\psi\rangle\right|^2 \\ &\geq 1 - \frac{2}{l+1}.\end{aligned}$$

This implies that

$$\big\| |\Psi_A\rangle - |\Psi_B\rangle \big\| \leq \frac{2}{\sqrt{l+1}}.$$

Let $|\Psi_A^q\rangle$ and $|\Psi_B^q\rangle$ be the final states of algorithm $\mathcal{A}$ and $\mathcal{B}$ before measurement respectively. Then by induction on the number of queries, we have

$$\big\| |\Psi_A^q\rangle - |\Psi_B^q\rangle \big\| \leq \frac{2q}{\sqrt{l+1}} .$$

This concludes the proof by noticing that

$$\mathrm{TD}\big(|\Psi_A^q\rangle, |\Psi_B^q\rangle\big) \leq \big\| |\Psi_A^q\rangle - |\Psi_B^q\rangle \big\| .$$

Finally, we show that if $\mathcal{A}$ is polynomial-time, then so is $\mathcal{B}$. Based on the construction of $\mathcal{B}$, it suffices to show that the reflection R is efficiently implementable for any polynomially large l. Here we use a result by Barenco et al. [8] which provides an efficient implementation for the projection Π onto $\vee^{l+1}\mathcal{H}$. More

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

precisely, they design a quantum circuit of size $O(\mathrm{poly}(l, \log \dim \mathcal{H}))$ that implements a unitary U such that $U|\phi\rangle = \sum_j |\xi_j\rangle|j\rangle$ on $\mathcal{H}^{\otimes(l+1)} \otimes \mathcal{H}'$ for an auxiliary space $\mathcal{H}'$ of dimension $O(l!)$. Here $|\xi_0\rangle = \Pi|\phi\rangle$ corresponds to the projection of $|\phi\rangle$ on the symmetric subspace. With U, we can implement the reflection R as $U^* S U$ where S is the unitary that introduces a minus sign conditioned on the second register being 0.

$$S|\Psi\rangle|j\rangle = \begin{cases} -|\Psi\rangle|j\rangle & \text{if } j = 0, \\ |\Psi\rangle|j\rangle & \text{otherwise.} \end{cases}$$

4.2 Quantum Money from PRS

Using Theorem 3, we can improve Theorem 2 to the following version. The proof is omitted as it is very similar to that for Theorem 2 and uses the complexity-theoretic no-cloning theorem [1,2] for Haar random states.

Theorem 5 (Cryptographic no-cloning Theorem with Oracle). *For any PRS $\{|\phi_k\rangle\}_{k\in\mathcal{K}}$, $m \in \mathrm{poly}(\kappa)$, $m < m'$ and any polynomial-time quantum query algorithm $\mathcal{C}$, the success cloning probability*

$$\mathop{\mathbb{E}}_{k\in\mathcal{K}} \left\langle \left(|\phi_k\rangle\langle\phi_k|\right)^{\otimes m'}, \mathcal{C}^{O_{\phi_k}}\left(\left(|\phi_k\rangle\langle\phi_k|\right)^{\otimes m}\right) \right\rangle = \mathrm{negl}(\kappa).$$

A direct application of this no-cloning theorem is that it gives rise to new constructions for private-key quantum money. As one of the earliest findings in quantum information [62,9], quantum money schemes have received revived interests in the past decade (see e.g. [1,40,43,21,22,3]). First, we recall the definition of quantum money scheme adapted from [2].

Definition 4 (Quantum Money Scheme). *A private-key quantum money scheme $\mathcal{S}$ consists of three algorithms:*

- *KeyGen, which takes as input the security parameter 1^κ and randomly samples a private key k.*
- *Bank, which takes as input the private key k and generates a quantum state $|\$\rangle$ called a **banknote**.*
- *Ver, which takes as input the private key k and an alleged banknote $|\dot{\varsigma}\rangle$, and either accepts or rejects.*

*The money scheme $\mathcal{S}$ has **completeness error** ε if $\mathsf{Ver}(k, |\$\rangle)$ accepts with probability at least $1 - \varepsilon$ for all valid banknote $|\$\rangle$.*

*Let Count be the money counter that output the number of valid banknotes when given a collection of (possibly entangled) alleged banknotes $|\dot{\varsigma}_1, \dot{\varsigma}_2, \ldots, \dot{\varsigma}_r\rangle$. Namely, Count will call Ver on each banknotes and return the number of times that Ver accepts. The money scheme $\mathcal{S}$ has **soundness error** δ if for any polynomial-time counterfeiter C that maps q valid banknotes $|\$_1\rangle, \ldots, |\$_q\rangle$ to r alleged banknotes $|\dot{\varsigma}_1, \ldots, \dot{\varsigma}_r\rangle$ satisfies*

$$\Pr\left[\mathsf{Count}\left(k, C(|\$_1\rangle, \ldots, |\$_q\rangle)\right) > q\right] \leq \delta.$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

The scheme $\mathcal{S}$ is ***secure*** *if it has completeness error $\leq 1/3$ and negligible soundness error.*

For any $\mathsf{PRS} = \{|\phi_k\rangle\}_{k\in\mathcal{K}}$ with key space $\mathcal{K}$, we can define a private-key quantum money scheme $\mathcal{S}_{\mathsf{PRS}}$ as follows:

- $\mathsf{KeyGen}(1^\kappa)$ randomly outputs $k \in \mathcal{K}$.
- $\mathsf{Bank}(k)$ generates the banknote $|\$\rangle = |\phi_k\rangle$.
- $\mathsf{Ver}(k, \rho)$ applies the projective measurement that accepts ρ with probability $\langle\phi_k|\rho|\phi_k\rangle$.

We remark that usually the money state $|\$\rangle$ takes the form $|\$\rangle = |s, \psi_s\rangle$ where the first register contains a classical serial number. Our scheme, however, does not require the use of the serial numbers. This simplification is brought to us by the strong requirement that any polynomial copies of $|\phi_k\rangle$ are indistinguishable from Haar random states.

Theorem 6. *The private-key quantum money scheme $\mathcal{S}_{PRS}$ is secure for all PRS.*

Proof. It suffices to prove the soundness of S_{PRS} is negligible. Assume to the contrary that there is a counterfeiter C such that

$$\Pr\big[\mathsf{Count}\big(k, C(|\phi_k\rangle^{\otimes q})\big) > q\big] \geq \kappa^{-c}$$

for some constant $c > 0$ and sufficiently large κ. From the counterfeiter C, we will construct an oracle algorithm $\mathcal{A}^{O_{\phi_k}}$ that maps q copies of $|\phi_k\rangle$ to $q+1$ copies with noticeable probability and this leads to a contradiction with Theorem 5.

The oracle algorithm $\mathcal{A}$ first runs C and implement the measurement

$$\Big\{M^0 = \mathbb{1} - |\phi_k\rangle\langle\phi_k|, M^1 = |\phi_k\rangle\langle\phi_k|\Big\}$$

on each copy of the money state C outputs. This measurement can be implemented by attaching an auxiliary qubit initialized in $(|0\rangle + |1\rangle)/\sqrt{2}$ and call the reflection oracle O_ϕ conditioned on the qubit being at 1 and performs the X measurement on this auxiliary qubit. This gives r-bit of outcome $\mathbf{x} \in \{0,1\}^r$. If $\mathbf{x}$ has Hamming weight at least $q+1$, algorithm $\mathcal{A}$ outputs any $q+1$ registers that corresponds to outcome 1; otherwise, it outputs $|0\rangle^{\otimes(q+1)}$. By the construction of $\mathcal{A}$, it succeeds in cloning $q+1$ money states from q copies with probability at least κ^{-c}.

Our security proof of the quantum money scheme is arguably simpler than that in [2]. In [2], to prove their hidden subspace money scheme is secure, one needs to develop the so called inner-product adversary method to show the worst-case query complexity for the hidden subspace states and use a random self-reducible argument to establish the average-case query complexity. In our case, it follows almost directly from the cryptographic no-cloning theorem with oracles. The quantum money schemes derived from PRS's enjoy many nice features of the hidden subspace scheme. Most importantly, they are also *query-secure* [2],

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

meaning that the bank can simply return the money state back to the user after verification.

It is also interesting to point out that quantum money states are not necessarily pseudorandom states. The hidden subspace state [2], for example, do not satisfy our definition of PRS as one can measure polynomially many copies of the state in the computational basis and recover a basis for the hidden subspace with high probability.

5 Entanglement of Pseudorandom Quantum States

In this section, we study the entanglement property of pseudorandom quantum states. Our result shows that any PRS consists of states that have high entanglement on average.

The entanglement property of a bipartite pure quantum state is well understood and is completely determined by the Schmidt coefficients of a bipartite state (see e.g. [32]). Any state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ on system A and B can be written as

$$|\psi\rangle = \sum_{j=1}^{R} \sqrt{\lambda_j} |\psi_A^j\rangle \otimes |\psi_B^j\rangle,$$

where $\lambda_j > 0$ for all $1 \leq j \leq R$ and the states $|\psi_A^j\rangle$ (and $|\psi_B^j\rangle$) form a set of orthonormal states on A (and B respectively). Here, the positive real numbers λ_j's are the Schmidt coefficients and R is the Schmidt rank of state $|\psi\rangle$. Let ρ_A be the reduced density matrix of $|\psi\rangle$ on system A, then λ_j is the nonzero eigenvalues of ρ_A. Entanglement can be measured by the Schmidt rank R or entropy-like quantities derived from the Schmidt coefficients. We consider the quantum α-Rényi entropy of ρ_A

$$S_\alpha(\rho_A) := \frac{1}{1-\alpha} \log\left(\sum_{j=1}^{R} \lambda_j^\alpha\right).$$

When $\alpha \to 1$, S_α coincides with the von Neumann entropy of ρ_A

$$S(\rho_A) = -\sum_{j=1}^{R} \lambda_j \log \lambda_j.$$

When $\alpha \to \infty$, S_α coincides with the quantum min entropy of ρ_A

$$S_{\min}(\rho_A) = -\log \|\rho_A\| = -\log \lambda_{\max},$$

where $\lambda_{\max}$ is the largest eigenvalue of ρ_A. For $\alpha = 2$, the entropy S_2 is the quantum analogue of the collision entropy.

For Haar random state $|\psi\rangle \sim \mu(\mathcal{H}_A \otimes \mathcal{H}_B)$ where the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$ are d_A and d_B respectively, the Page conjecture [50] proved in [23,55,56] states that for $d_A \leq d_B$, the average entanglement entropy is explicitly given as

$$\mathbb{E}\, S(\rho_A) = \frac{1}{\ln 2}\left[\left(\sum_{j=d_B+1}^{d_A d_B} \frac{1}{j}\right) - \frac{d_B - 1}{2d_A}\right] > \log d_A - O(1).$$

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

That is, the Haar random states are highly entangled on average and, in fact, a typical Haar random state is almost maximumly entangled. A more detailed discussion on this phenomena is give in [29,35]. The following theorem and its corollary tell us that pseudorandom states are also entangled on average though the quantitative bound is much weaker.

Theorem 7. *Let $\{|\phi_k\rangle\}_{k\in\mathcal{K}}$ be a family of PRS with security parameter κ. Consider partitions of the state $|\phi_k\rangle$ into systems A and B consisting of n_A and n_B qubits each where both n_A and n_B are polynomial in the security parameter. Let ρ_k be the reduced density matrix on system A. Then,*

$$\mathop{\mathbb{E}}_k \operatorname{tr}(\rho_k^2) = \operatorname{negl}(\kappa).$$

Proof. Assume to the contrary that

$$\mathop{\mathbb{E}}_k \operatorname{tr}(\rho_k^2) \geq \kappa^{-c}$$

for some constant $c > 0$ and sufficiently large κ. We will construct a distinguisher $\mathcal{A}$ that tells the family of state $\{|\phi_k\rangle\}$ apart from the Haar random states.

Consider the SWAP test performed on the system A of two copies of $|\phi_k\rangle$. The test accepts with probability

$$\frac{1+\operatorname{tr}(\rho_k^2)}{2}.$$

Let distinguisher $\mathcal{A}$ be the above SWAP test, we have

$$\begin{aligned}&\left|\Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}(|\phi_k\rangle^{\otimes 2})=1\right]-\Pr_{|\psi\rangle\leftarrow\mu}\left[\mathcal{A}(|\psi\rangle^{\otimes 2})=1\right]\right|\\ =&\frac{1}{2}\left|\mathop{\mathbb{E}}_k\operatorname{tr}(\rho_k^2)-\mathop{\mathbb{E}}_\mu\operatorname{tr}(\rho_\psi^2)\right|\geq\kappa^{-c}/4,\end{aligned}$$

for sufficiently large κ. The last step follows by a formula of Lubkin [37]

$$\mathop{\mathbb{E}}_{|\psi\rangle\leftarrow\mu}\operatorname{tr}(\rho_\psi^2)=\frac{d_A+d_B}{d_Ad_B+1}=\frac{2^{n_A}+2^{n_B}}{2^{n_A+n_B}+1}=\operatorname{negl}(\kappa).$$

Corollary 1. *Let $\{|\phi_k\rangle\}_{k\in\mathcal{K}}$ be a family of PRS with security parameter κ. Consider partitions of the state $|\phi_k\rangle$ into systems A and B consisting of n_A and n_B qubits each where both n_A and n_B are polynomial in the security parameter. We have*

1. *Let R_k be the Schmidt rank of state $|\phi_k\rangle$ under the A, B partition, then $\mathbb{E}_k R_k \geq \kappa^c$ for all constant $c > 0$ and sufficiently large κ.*
2. *$\mathbb{E}_k S_{\min}(\rho_k) = \omega(\log\kappa)$ and $\mathbb{E}_k S(\rho_k) = \omega(\log\kappa)$.*

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

Proof. The first item follows from the fact that

$$\mathrm{tr}(\rho_k^2) \geq \frac{1}{R_k}.$$

where R_k is the Schmidt rank of state $|\phi_k\rangle$. The second item for the min entropy follows by Jensen's inequality and

$$\mathrm{tr}(\rho_k^2) \geq \lambda_{\max}^2.$$

Finally, the bound on the expected entanglement entropy follows by the fact that min entropy is the smallest α-Rényi entropy for all $\alpha > 0$.

6 Pseudorandom Unitary Operators (PRUs)

6.1 Definitions

Our notion of pseudorandom states readily extends to distributions over unitary operators. Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K}$ a key space, both of which depend on a security parameter κ. Let μ be the Haar measure on the unitary group $\mathrm{U}(\mathcal{H})$.

Definition 5. *A family of unitary operators $\{U_k \in \mathrm{U}(\mathcal{H})\}_{k\in\mathcal{K}}$ is **pseudorandom**, if two conditions hold:*

1. *(**Efficient computation**) There is an efficient quantum algorithm Q, such that for all k and any $|\psi\rangle \in \mathrm{S}(\mathcal{H})$, $Q(k, |\psi\rangle) = U_k|\psi\rangle$.*
2. *(**Pseudorandomness**) U_k with a random key k is **computationally indistinguishable** from a Haar random unitary operator. More precisely, for any efficient quantum algorithm $\mathcal{A}$ that makes at most polynomially many queries to the oracle,*

$$\left| \Pr_{k\leftarrow\mathcal{K}}\left[\mathcal{A}^{U_k}(1^\kappa) = 1\right] - \Pr_{U\leftarrow\mu}\left[\mathcal{A}^{U}(1^\kappa) = 1\right] \right| = \mathrm{negl}(\kappa).$$

The extensive literature on approximation of Haar randomness on unitary groups concerns with unitary *designs* [19,12], which are statistical approximations to the Haar random distribution up to a fixed t-th moment. Our notion of pseudorandom unitary operators in terms of computational indistinguishability, in addition to independent interest, supplements and could substitute for unitary designs in various applications.

6.2 Candidate constructions

Clearly, given a pseudorandom unitary family $\{U_k\}$, it immediately gives pseudorandom states as well (e.g., $\{U_k|0\rangle\}$). On the other hand, our techniques for constructing pseudorandom states can be extended to give candidate constructions for pseudorandom unitary operators (PRUs) in the following way. Let

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

$\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$. Assume we have a pseudorandom function $\mathsf{PRF} : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$, with domain $\mathcal{X} = \{0, 1, 2, \ldots, N-1\}$ and $N = 2^n$. Using the phase kick-back technique, we can implement the unitary transformation $T_k \in \mathrm{U}(\mathcal{H})$ that maps

$$T_k : |x\rangle \mapsto \omega_N^{\mathsf{PRF}_k(x)}|x\rangle, \quad \omega_N = \exp(2\pi i/N). \tag{9}$$

Our pseudorandom states were given by $|\phi_k\rangle = T_k H^{\otimes n}|0\rangle$, where $H^{\otimes n}$ denotes the n-qubit Hadamard transform. We conjecture that by repeating the operation $T_k H^{\otimes n}$ a constant number of times (with different keys k), we get a PRU. This is resembles the construction of unitary t-designs in [45,44].

Alternatively, one can give a candidate construction for PRUs based on pseudorandom permutations (PRPs) as follows. First, let PRP_k be a pseudorandom permutation (with key $k \in \mathcal{K}$) acting on $\{0,1\}^n$, and suppose we have efficient quantum circuits that compute the permutation $P_k : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus \mathsf{PRP}_k(x)\rangle$ as well as its inverse $R_k : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus \mathsf{PRP}_k^{-1}(x)\rangle$ (where $\oplus$ denotes the bitwise xor operation). Then we can compute the permutation in-place by applying the following sequence of operations:

$$\begin{aligned} |x\rangle|0\rangle &\xrightarrow{P_k} |x\rangle|\mathsf{PRP}_k(x)\rangle \\ &\xrightarrow{SWAP} |\mathsf{PRP}_k(x)\rangle|x\rangle \\ &\xrightarrow{R_k} |\mathsf{PRP}_k(x)\rangle|0\rangle. \end{aligned} \tag{10}$$

For simplicity, let us denote this operation by $S_k : |x\rangle \mapsto |\mathsf{PRP}_k(x)\rangle$ (ignoring the second register, which stays in the state $|0\rangle$). Now we can consider repeating the operation $S_k H^{\otimes n}$ several times (with different keys k), as a candidate for a PRU. Note that this resembles the construction of unitary t-designs in [27].

It is an interesting challenge to prove that these constructions actually yield PRUs. For the special case of non-adaptive adversaries, one could try to use the proof techniques of [27,45,44] for unitary t-designs. For the general case, where the adversary can make adaptive queries to the pseudorandom unitary, new proof techniques seem to be needed. Finally, we can consider combining all of these ingredients (the pseudorandom operations S_k and T_k, and the Hadamard transform) to try to obtain more efficient constructions of PRUs.

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

References

1. Aaronson, S.: Quantum copy-protection and quantum money. In: Proceedings of the Twenty-Fourth Annual IEEE Conference on Computational Complexity (CCC'09). pp. 229–242. IEEE Computer Society (2009), http://dx.doi.org/10.1109/CCC.2009.42
2. Aaronson, S., Christiano, P.: Quantum money from hidden subspaces. In: Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing. pp. 41–60. STOC '12, ACM, New York, NY, USA (2012), http://doi.acm.org/10.1145/2213977.2213983
3. Aaronson, S., Farhi, E., Gosset, D., Hassidim, A., Kelner, J., Lutomirski, A.: Quantum money. Commun. ACM 55(8), 84–92 (Aug 2012), http://doi.acm.org/10.1145/2240236.2240258
4. Ambainis, A., Emerson, J.: Quantum t-designs: t-wise Independence in the Quantum World. In: Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07). pp. 129–140 (June 2007)
5. Ambainis, A., Rosmanis, A., Unruh, D.: Quantum attacks on classical proof systems: The hardness of quantum rewinding. In: Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science. pp. 474–483. IEEE Computer Society (2014), http://dx.doi.org/10.1109/FOCS.2014.57, full version at https://arxiv.org/abs/1404.6898
6. Banerjee, A., Peikert, C., Rosen, A.: Pseudorandom functions and lattices. In: Advances in Cryptology–Eurocrypt 2012. pp. 719–737. Springer-Verlag (2012), http://dx.doi.org/10.1007/978-3-642-29011-4_42
7. Barak, B., Shaltiel, R., Wigderson, A.: Computational analogues of entropy. In: Arora, S., Jansen, K., Rolim, J.D.P., Sahai, A. (eds.) Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques. pp. 200–215. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
8. Barenco, A., Berthiaume, A., Deutsch, D., Ekert, A., Jozsa, R., Macchiavello, C.: Stabilization of quantum computations by symmetrization. SIAM Journal on Computing 26(5), 1541–1557 (1997), https://doi.org/10.1137/S0097539796302452
9. Bennett, C.H., Brassard, G., Breidbart, S., Wiesner, S.: Quantum Cryptography, or Unforgeable Subway Tokens, pp. 267–275. Springer, Boston, MA (1983)
10. Blum, M., Micali, S.: How to Generate Cryptographically Strong Sequences of Pseudorandom Bits. SIAM Journal on Computing 13(4), 850–864 (1984), https://doi.org/10.1137/0213053
11. Brandão, F.G.S.L., Harrow, A.W., Horodecki, M.: Efficient quantum pseudorandomness. Phys. Rev. Lett. 116, 170502 (Apr 2016), https://link.aps.org/doi/10.1103/PhysRevLett.116.170502
12. Brandão, F.G.S.L., Harrow, A.W., Horodecki, M.: Local random quantum circuits are approximate polynomial-designs. Communications in Mathematical Physics 346(2), 397–434 (Sep 2016), https://doi.org/10.1007/s00220-016-2706-8
13. Bremner, M.J., Mora, C., Winter, A.: Are random pure states useful for quantum computation? Phys. Rev. Lett. 102, 190502 (May 2009), https://link.aps.org/doi/10.1103/PhysRevLett.102.190502
14. Chen, Y.H., Chung, K.M., Lai, C.Y., Vadhan, S.P., Wu, X.: Computational notions of quantum min-entropy. arXiv:1704.07309 (2017)
15. Chung, K.M., Shi, Y., Wu, X.: Physical randomness extractors: generating random numbers with minimal assumptions. arXiv preprint arXiv:1402.4797 (2014)

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

16. Cleve, R., Leung, D., Liu, L., Wang, C.: Near-linear constructions of exact unitary 2-designs. Quantum Information & Computation 16(9&10), 721–756 (2016), http://www.rintonpress.com/xxqic16/qic-16-910/0721-0756.pdf
17. Dankert, C., Cleve, R., Emerson, J., Livine, E.: Exact and approximate unitary 2-designs and their application to fidelity estimation. Phys. Rev. A 80, 012304 (Jul 2009), https://link.aps.org/doi/10.1103/PhysRevA.80.012304
18. Dieks, D.: Communication by EPR devices. Physics Letters A 92(6), 271–272 (1982)
19. Emerson, J., Weinstein, Y.S., Saraceno, M., Lloyd, S., Cory, D.G.: Pseudo-random unitary operators for quantum information processing. science 302(5653), 2098–2100 (2003)
20. Farhi, E., Gosset, D., Hassidim, A., Lutomirski, A., Nagaj, D., Shor, P.: Quantum state restoration and single-copy tomography for ground states of hamiltonians. Phys. Rev. Lett. 105, 190503 (Nov 2010), https://link.aps.org/doi/10.1103/PhysRevLett.105.190503
21. Farhi, E., Gosset, D., Hassidim, A., Lutomirski, A., Nagaj, D., Shor, P.: Quantum State Restoration and Single-Copy Tomography for Ground States of Hamiltonians. Phys. Rev. Lett. 105, 190503 (Nov 2010), https://link.aps.org/doi/10.1103/PhysRevLett.105.190503
22. Farhi, E., Gosset, D., Hassidim, A., Lutomirski, A., Shor, P.: Quantum money from knots. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. pp. 276–289. ITCS '12, ACM, New York, NY, USA (2012), http://doi.acm.org/10.1145/2090236.2090260
23. Foong, S.K., Kanno, S.: Proof of page's conjecture on the average entropy of a subsystem. Phys. Rev. Lett. 72, 1148–1151 (Feb 1994), https://link.aps.org/doi/10.1103/PhysRevLett.72.1148
24. Goldreich, O., Goldwasser, S., Micali, S.: On the cryptographic applications of random functions. In: Advances in Cryptology – CRYPTO 1984. pp. 276–288. Springer-Verlag New York, Inc. (1985)
25. Goldreich, O., Goldwasser, S., Micali, S.: How to Construct Random Functions. J. ACM 33(4), 792–807 (Aug 1986), http://doi.acm.org/10.1145/6490.6503
26. Harrow, A.W.: The church of the symmetric subspace. arXiv:1308.6595 (2013)
27. Harrow, A.W., Low, R.A.: Efficient quantum tensor product expanders and k-designs. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pp. 548–561. Springer (2009)
28. Håstad, J., Impagliazzo, R., Levin, L.A., Luby, M.: A pseudorandom generator from any one-way function. SIAM Journal on Computing 28(4), 1364–1396 (1999)
29. Hayden, P., Leung, D.W., Winter, A.: Aspects of generic entanglement. Communications in Mathematical Physics 265(1), 95–117 (Jul 2006), https://doi.org/10.1007/s00220-006-1535-6
30. Helstrom, C.W.: Detection theory and quantum mechanics. Information and Control 10(3), 254–291 (1967)
31. Holevo, A.S.: An analogue of statistical decision theory and noncommutative probability theory. Trudy Moskovskogo Matematicheskogo Obshchestva 26, 133–149 (1972)
32. Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K.: Quantum entanglement. Rev. Mod. Phys. 81, 865–942 (Jun 2009), https://link.aps.org/doi/10.1103/RevModPhys.81.865
33. Impagliazzo, R., Wigderson, A.: P = BPP if E Requires Exponential Circuits: Derandomizing the XOR Lemma. In: Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing. pp. 220–229. STOC '97, ACM, New York, NY, USA (1997), http://doi.acm.org/10.1145/258533.258590

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

34. Kueng, R., Gross, D.: Qubit stabilizer states are complex projective 3-designs (2015), arXiv:1510.02767
35. Liu, Z.W., Lloyd, S., Zhu, E.Y., Zhu, H.: Entropic scrambling complexities. arXiv:1703.08104 (2017)
36. Low, R.A.: Large deviation bounds for k-designs. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 465(2111), 3289–3308 (2009), http://rspa.royalsocietypublishing.org/content/465/2111/3289
37. Lubkin, E.: Entropy of an n-system from its correlation with ak-reservoir. Journal of Mathematical Physics 19(5), 1028–1031 (1978)
38. Luby, M., Rackoff, C.: How to construct pseudorandom permutations from pseudorandom functions. SIAM Journal on Computing 17(2), 373–386 (1988)
39. Lutomirski, A.: An online attack against wiesner's quantum money (2010), arXiv:1010.0256
40. Lutomirski, A., Aaronson, S., Farhi, E., Gosset, D., Hassidim, A., Kelner, J., Shor, P.: Breaking and making quantum money: toward a new quantum cryptographic protocol. In: Proceedings of the Innovations in Theoretical Computer Science Conference. pp. 20–31. ITCS '10, Tsinghua University Press (2010)
41. Mezher, R., Ghalbouni, J., Dgheim, J., Markham, D.: Efficient quantum pseudorandomness with simple graph states (2017), arXiv:1709.08091
42. Miller, C.A., Shi, Y.: Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. Journal of the ACM (JACM) 63(4), 33 (2016)
43. Mosca, M., Stebila, D.: Quantum coins. In: Bruen, A.A., Wehlau, D.L. (eds.) Error-Correcting Codes, Finite Geometries and Cryptography. Contemporary Mathematics, vol. 523, pp. 35–47. American Mathematical Society (2010), http://www.ams.org/bookstore?fn=20&arg1=conmseries&ikey=CONM-523
44. Nakata, Y., Hirche, C., Koashi, M., Winter, A.: Efficient quantum pseudorandomness with nearly time-independent hamiltonian dynamics. Phys. Rev. X 7, 021006 (Apr 2017), https://link.aps.org/doi/10.1103/PhysRevX.7.021006
45. Nakata, Y., Hirche, C., Morgan, C., Winter, A.: Unitary 2-designs from random x- and z-diagonal unitaries. Journal of Mathematical Physics 58(5), 052203 (2017), https://doi.org/10.1063/1.4983266
46. Nakata, Y., Koashi, M., Murao, M.: Generating a state t -design by diagonal quantum circuits. New Journal of Physics 16(5), 053043 (2014), http://stacks.iop.org/1367-2630/16/i=5/a=053043
47. Naor, M., Reingold, O.: Synthesizers and their application to the parallel construction of pseudo-random functions. J. Comput. Syst. Sci. 58(2), 336–375 (Apr 1999), http://dx.doi.org/10.1006/jcss.1998.1618
48. Nisan, N., Wigderson, A.: Hardness vs randomness. J. Comput. Syst. Sci. 49(2), 149–167 (Oct 1994), http://dx.doi.org/10.1016/S0022-0000(05)80043-1
49. Ortigoso, J.: Twelve years before the quantum no-cloning theorem. arXiv:1707.06910 (2017)
50. Page, D.N.: Average entropy of a subsystem. Phys. Rev. Lett. 71, 1291–1294 (Aug 1993), https://link.aps.org/doi/10.1103/PhysRevLett.71.1291
51. Park, J.L.: The concept of transition in quantum mechanics. Found. Phys. 1, 23–33 (1970)
52. Popescu, S., Short, A.J., Winter, A.: Entanglement and the foundations of statistical mechanics. Nature Physics 2(11), 754 (2006)
53. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. Journal of the ACM (JACM) 56(6), 34 (2009)

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

54. Rompel, J.: One-way functions are necessary and sufficient for secure signatures. In: Proceedings of the twenty-second annual ACM symposium on Theory of computing. pp. 387–394. ACM (1990)
55. Sánchez-Ruiz, J.: Simple proof of page's conjecture on the average entropy of a subsystem. Phys. Rev. E 52, 5653–5655 (Nov 1995), https://link.aps.org/doi/10.1103/PhysRevE.52.5653
56. Sen, S.: Average entropy of a quantum subsystem. Phys. Rev. Lett. 77, 1–3 (Jul 1996), https://link.aps.org/doi/10.1103/PhysRevLett.77.1
57. Shamir, A.: On the generation of cryptographically strong pseudorandom sequences. ACM Trans. Comput. Syst. 1(1), 38–44 (Feb 1983), http://doi.acm.org/10.1145/357353.357357
58. Song, F.: Quantum-secure pseudorandom permutations (June, 2017), blog post, available at http://qcc.fangsong.info/2017-06-quantumprp/
59. Watrous, J.: The theory of quantum information. To be published by Cambridge University Press (2018), a draft copy is available at https://cs.uwaterloo.ca/~watrous/TQI/
60. Webb, Z.: The clifford group forms a unitary 3-design. Quantum Information & Computation 16(15&16), 1379–1400 (2016), http://www.rintonpress.com/xxqic16/qic-16-1516/1379-1400.pdf
61. Werner, R.F.: Optimal cloning of pure states. Phys. Rev. A 58, 1827–1832 (Sep 1998), https://link.aps.org/doi/10.1103/PhysRevA.58.1827
62. Wiesner, S.: Conjugate Coding. SIGACT News 15(1), 78–88 (1983), original manuscript written circa 1970
63. Wootters, W.K., Zurek, W.H.: A single quantum cannot be cloned. Nature 299, 802–803 (Oct 1982)
64. Yao, A.C.: Theory and application of trapdoor functions. In: 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982). pp. 80–91 (Nov 1982)
65. Yuen, H.: A quantum lower bound for distinguishing random functions from random permutations. Quantum Information & Computation 14(13-14), 1089–1097 (2014), http://dl.acm.org/citation.cfm?id=2685166
66. Zhandry, M.: How to Construct Quantum Random Functions. In: FOCS 2012. pp. 679–687. IEEE (2012), http://eprint.iacr.org/2012/182
67. Zhandry, M.: A Note on the Quantum Collision and Set Equality Problems. Quantum Information and Computation 15(7 & 8) (2015), http://arxiv.org/abs/1312.1027
68. Zhandry, M.: A Note on Quantum-Secure PRPs (2016), available at https://eprint.iacr.org/2016/1076
69. Zhandry, M.: Quantum lightning never strikes the same state twice (2017), iACR eprint 2017/1080
70. Zhu, H.: Multiqubit clifford groups are unitary 3-designs (2015), arXiv:1510.02619

Liu, Yi-Kai; Ji, Zhengfeng; Song, Fang.
"Pseudorandom Quantum States."
Paper presented at Crypto 2018 - 38th International Cryptology Conference, Santa Barbara, CA, United States. August 19, 2018 - August 23, 2018.

Exact Moments of Mutual Information of Jacobi MIMO Channels in High-SNR Regime

Lu Wei
Department of Electrical and Computer Engineering
University of Michigan-Dearborn
Dearborn, MI 48128, USA
Email: luwe@umich.edu

Zhong Zheng and Hamid Gharavi
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899, USA
Emails: {zhong.zheng, hamid.gharavi}@nist.gov

***Abstract*—In this paper, we propose an analytical framework to derive all positive integer moments of MIMO mutual information in high-SNR regime. The approach is based on efficient use of the underlying matrix integrals of the high-SNR mutual information. As an example, the framework is applied to the study of Jacobi MIMO channel model relevant to fiber optical and interference-limited multiuser MIMO communications. For such a channel model, we obtain explicit expressions for the exact moments of the mutual information in the high-SNR regime. The derived moments are utilized to construct approximations to the corresponding outage probability. Simulation shows the usefulness of the results in a crucial scenario of low outage probability with finite number of antennas.**

I. Introduction

Mutual information is among the most important quantities in information theory. For Multiple-Input-Multiple-Output (MIMO) communications, the supremum of mutual information provides the fundamental performance measure, the channel capacity. A great effort has been made to understand the statistical behavior of MIMO mutual information of various channel models. Existing knowledge in the literature is, however, mostly limited to either exact mean values (first moments) [1–4] or asymptotic (in channel dimensions) means and variances [5, 6]. The first moment corresponds to the ergodic mutual information, whereas the higher moments are needed to describe the outage probability relevant to slow or block fading scenarios. Another motivation of our study is that the prevailingly adopted asymptotic variances based approximate outage probabilities [5, 6] fail to capture the true one when the number of antennas is small and/or the outage probability is low. An accurate characterization requires the exact higher moments of mutual information that governs the tail behavior of the distribution.

Determining the exact higher moments of MIMO mutual information for an arbitrary Signal-to-Noise Ratio (SNR) is a well-known difficult task. We will show in this paper that progress can be made when assuming the SNR is high. In particular, we propose an analytical framework to obtain the exact moments of any order in the high-SNR regime, which is valid for a family of channel models. The idea comes from the observation that moments of high-SNR mutual information can be efficiently calculated via the underlying matrix integrals. The high-SNR regime provides crucial insights to the behavior of the MIMO channels. For example, it characterizes the minimum required transmit power, also known as the high-SNR power offset [7].

To demonstrate the usefulness of the proposed framework, we study the mutual information of the Jacobi MIMO channels. The Jacobi MIMO channel is a useful channel model for MIMO optical communications [2, 6] as well as the interference-limited multiuser MIMO [5]. The main result of this paper is the exact yet explicit expressions for the mutual information moments of any order of the Jacobi MIMO channels in the high-SNR regime.

II. Problem Formulation

A. MIMO Mutual Information

For a generic MIMO system consisting of n transmit and m receive antennas, the communication channel in between is described by an $m \times n$ random matrix $\mathbf{H}$. Assuming i.i.d. input across transmit antennas and that the channel matrix $\mathbf{H}$ is only known to the receiver, the mutual information in nats/second/Hz of the MIMO system is [1]

$$\mathrm{I} = \ln\det\left(\mathbf{I}_m + r\mathbf{H}\mathbf{H}^\dagger\right) = \sum_{i=1}^{m} \ln\left(1 + r\theta_i\right), \qquad (1)$$

where it can be made without loss of generality to first assume that $m \le n$. Here, $\ln(\cdot)$ is the natural logarithm, $\det(\cdot)$ is the matrix determinant, r is the SNR, and $\theta_m \le \cdots \le \theta_2 \le \theta_1$ denote the eigenvalues of the Hermitian matrix $\mathbf{H}\mathbf{H}^\dagger$. In the high-SNR regime, by ignoring the constant $\mathbf{I}_m$ in (1) the mutual information can be approximated by

$$\mathcal{I} = m\ln r + \ln\det\left(\mathbf{H}\mathbf{H}^\dagger\right) \qquad (2a)$$

$$= m\ln r + \sum_{i=1}^{m} \ln\theta_i. \qquad (2b)$$

The above approximation becomes exact as the SNR r grows to infinity.

A fundamental information-theoretic quantity of MIMO channels is outage probability, which is the probability that a given rate exceeds the value of the mutual information. For the high-SNR case (2), the channel outage probability $P_{\text{out}}(z)$ as a function of the rate z is defined as

$$P_{\text{out}}(z) = \mathbb{P}\left(\mathcal{I} < z\right). \qquad (3)$$

Gharavi, Hamid.
"Exact Moments of Mutual Information of Jacobi MIMO Channels in High-SNR Regime."
Paper presented at IEEE Global Communications Conference 2018 (GLOBECOM 2018), Abu Dhabi, United Arab Emirates. December 9, 2018 - December 13, 2018.

B. Jacobi MIMO Channels

As will be seen, the Jacobi MIMO channel is a channel model for both MIMO optical communications [2, 6] and interference-limited multiuser MIMO [5]. However, we will formulate the problem and set up the notations mainly in the context of the former application. The relevance to the latter application will only be briefly mentioned.

The spatial degrees of freedom of the MIMO Rayleigh channels provide the well-known linear capacity scaling law [1] with respect to the number of transceiver antennas. The idea of the MIMO fiber optical channels is to achieve a similar scaling law by also exploiting the spatial degrees of freedom. In particular, multiple spatial transmission within the same fiber is achieved by designing a multi-mode and/or multi-core fiber. As a first step to exploring the spatial diversity, the Jacobi MIMO optical channel has been proposed in [2, 6], which is based on the following assumptions. The propagation through the fiber is considered as lossless such that it is modeled as an $l \times l$ random unitary matrix $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}_l$, which is also known as the scattering matrix. Assuming n transmitting and m receiving modes with $m \leq n$, the effective MIMO optical channel[1] $\mathbf{H}$ equals the upper left sub-matrix of the scattering matrix $\mathbf{U} = (u_{ij})$ with the condition $l > m + n$, i.e.,

$$\mathbf{H} = (u_{ij})_{i=1,\ldots,m;\ j=1,\ldots,n}\,. \tag{4}$$

Under the above assumptions, the joint eigenvalue density of the hermitianized channel matrix $\mathbf{H}\mathbf{H}^\dagger$ is given by [2, 6]

$$p\left(\boldsymbol{\theta}\right) = \frac{1}{c} \prod_{1 \leq i < j \leq m} \left(\theta_i - \theta_j\right)^2 \prod_{i=1}^{m} \theta_i^{\alpha_1} \left(1 - \theta_i\right)^{\alpha_2}, \tag{5}$$

where $0 \leq \theta_m \leq \cdots \leq \theta_2 \leq \theta_1 \leq 1$ and

$$\alpha_1 = n - m, \qquad \alpha_2 = l - m - n.$$

For the above parameters α_1 and α_2, the resulting normalization constant c is

$$c = \frac{\prod_{i=1}^{m} \Gamma(i+1)\Gamma(l-n-i+1)\Gamma(n-i+1)}{\Gamma(l-i+1)}.$$

The ensemble (5) is known as the Jacobi ensemble [8] in random matrix theory, and hence the name Jacobi MIMO channels in the communications theory/information theory community. The eigenvalue density of the interference-limited MIMO channel considered in [5] takes the form of (5) with the same parameter $\alpha_1 = n - m$ as the difference between the number of transmit and receive antennas. For this application, the parameter α_2 is now

$$\alpha_2 = kn - m,$$

where k is the number of interferers. For detailed information of the considered interference-limited MIMO channel including its connection to the Jacobi ensemble, we refer interested readers to [5].

[1]For a detailed physical interpretation of this channel model, we refer the readers to the excellent discussion in [6].

For the application to MIMO optical communications, the exact ergodic mutual information $\mathbb{E}\left[\mathrm{I}\right]$ of the Jacobi MIMO channels has been calculated in [2] by integrating the mutual information (1) over the eigenvalue density (5), whereas an unexplicit and an asymptotic second moment expressions are available in [6]. For the application to interference-limited MIMO channels, the first two asymptotic moments as well as a differential equation for the moments have been derived in [5]. The exact higher moments of the mutual information $\mathbb{E}\left[\mathrm{I}^k\right]$, $k = 2, 3, \ldots$, which are needed to characterize the outage probability, remain an open problem[2]. Despite the fact that even the exact second moment of the mutual information $\mathbb{E}\left[\mathrm{I}^2\right]$ is notoriously difficult to obtain, we will show that all the exact moments of the high-SNR mutual information (2), $\mathbb{E}\left[\mathcal{I}^k\right]$, $k = 1, 2, \ldots$, can be explicitly calculated. These moments are utilized to construct approximations to the outage probability in the high-SNR regime, which are in fact accurate for moderate SNR values as will be seen.

III. Exact Cumulants of High-SNR Mutual Information

To compute $\mathbb{E}\left[\mathcal{I}^k\right]$, one naturally would like to integrate (2b) over the eigenvalue density (5) rather than to integrate (2a) over the density of the matrix $\mathbf{H}\mathbf{H}^\dagger$. This is because the former integral only involves m variables, whereas the latter involves m^2 variables. Contrary to this intuition, we show that by working with the corresponding matrix integrals the exact moment of any order can be obtained in a straightforward manner. Instead of directly deriving the moments as in [7, 9], another ingredient that leads to our results is the study of the cumulants of $\mathcal{I}$, which is technically more convenient as will be seen.

Since the term $m \ln r$ in (2a) is a constant, we first focus on the statistics of the random variable

$$x = \ln\det\left(\mathbf{H}\mathbf{H}^\dagger\right). \tag{6}$$

The cumulant generating function $K(s)$ of x is defined as

$$K(s) = \ln \mathbb{E}\left[\mathrm{e}^{sx}\right] = \sum_{i=1}^{\infty} \tilde{\kappa}_i \frac{s^i}{i!}, \tag{7}$$

where the i-th cumulant $\tilde{\kappa}_i$ of x is recovered from the generating function as

$$\tilde{\kappa}_i = \frac{\mathrm{d}^i}{\mathrm{d}s^i} K(s)\bigg|_{s=0}. \tag{8}$$

Denoting the i-th cumulant of $\mathcal{I}$ by κ_i, we have

$$\kappa_1 = m \ln r + \tilde{\kappa}_1, \tag{9a}$$
$$\kappa_i = \tilde{\kappa}_i, \qquad i \geq 2, \tag{9b}$$

[2]A representation of the outage probability involving nested sums over partitions is also available in [6], which is computationally demanding and provides little insights.

Gharavi, Hamid.
"Exact Moments of Mutual Information of Jacobi MIMO Channels in High-SNR Regime."
Paper presented at IEEE Global Communications Conference 2018 (GLOBECOM 2018), Abu Dhabi, United Arab Emirates. December 9, 2018 - December 13, 2018.

which is obtained by the shift-equivariant and the shift-invariant property for cumulants, respectively. With the knowledge of cumulants of the mutual information $\mathcal{I}$, the corresponding moments can be determined. Specifically, the i-th moment of $\mathcal{I}$ is written in terms of the first i cumulants as

$$\mathbb{E}\left[\mathcal{I}^i\right] = B_i\left(\kappa_1, \ldots, \kappa_i\right), \tag{10}$$

where B_i is the Bell polynomial [10]. For example, the first five moments of $\mathcal{I}$ are listed below

$$\begin{aligned}
\mathbb{E}\left[\mathcal{I}\right] &= \kappa_1, && \text{(11a)}\\
\mathbb{E}\left[\mathcal{I}^2\right] &= \kappa_2 + \kappa_1^2, && \text{(11b)}\\
\mathbb{E}\left[\mathcal{I}^3\right] &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3, && \text{(11c)}\\
\mathbb{E}\left[\mathcal{I}^4\right] &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4, && \text{(11d)}\\
\mathbb{E}\left[\mathcal{I}^5\right] &= \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + \\
&\quad 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3 + \kappa_1^5. && \text{(11e)}
\end{aligned}$$

With the above preparation, we now present the main technical contribution of this paper.

Proposition 1. *The i-th exact cumulant κ_i of the high-SNR mutual information (2) of the Jacobi MIMO channels (4) is given by*

$$\begin{aligned}
\kappa_1 &= m\ln r + n\psi_0(n) - l\psi_0(l) - (n-m)\psi_0(n-m)\\
&\quad +(l-m)\psi_0(l-m), \quad i=1,\\
\kappa_i &= n\psi_{i-1}(n) - l\psi_{i-1}(l) - (n-m)\psi_{i-1}(n-m)\\
&\quad +(l-m)\psi_{i-1}(l-m) + (i-1)\big(\psi_{i-2}(n) -\\
&\quad \psi_{i-2}(l) - \psi_{i-2}(n-m) + \psi_{i-2}(l-m)\big), \quad i\geq 2,
\end{aligned}$$

where

$$\psi_i(z) = \frac{\partial^{i+1}\ln\Gamma(z)}{\partial z^{i+1}} = (-1)^{i+1}i!\sum_{k=0}^{\infty}\frac{1}{(k+z)^{i+1}} \tag{13}$$

are the polygamma functions [11].

The proof of Proposition 1 is in the Appendix. A similar, yet unsimplified, expression of κ_1 in Proposition 1 has been obtained in the context of mean power offset of interference-limited MIMO channels [7, Eq. (78)]. In such a setting, our derived higher cumulants will be useful to study the distribution of power offset for nonergodic channels.

Note that in the cases $i = 0, 1$, the polygamma functions (13) of positive integer argument can be reduced to finite sums as [11]

$$\begin{aligned}
\psi_0(l) &= -\gamma + \sum_{k=1}^{l-1}\frac{1}{k}, && \text{(14a)}\\
\psi_1(l) &= \frac{\pi^2}{6} - \sum_{k=1}^{l-1}\frac{1}{k^2}, && \text{(14b)}
\end{aligned}$$

which is known as the digamma function and the trigamma function, respectively, and $\gamma \approx 0.5772$ is Euler's constant.

It is worth mentioning that the proposed idea of using matrix-variate integrals to derive cumulants of MIMO mutual information can be equally applied to study other MIMO channel models. This includes the MIMO Rayleigh channels with[3] and without[4] antenna correlation as well as recent popular MIMO product channels [3, 4]. Results on all the integer moments of the above mentioned channel models in the high-SNR regime will be reported separately.

IV. Outage Probability in the High-SNR Regime

With the exact cumulants in Proposition 1 and the cumulant-moment relations (10), (11), closed-form moment-based approximations to the outage probability (3) can now be constructed. Moment-based approximation is a useful tool in situations when the exact distribution is intractable whereas the moments are available. The basic idea of moment-based approximation is to match the moments and support of an unknown distribution by an elementary distribution and the associated orthogonal polynomials [14, 15].

For convenience, in the simulation we construct the approximative outage probability via the moments of the random variable $-x$ in (6). Since $-x \in [0, \infty)$ has the same support as a gamma random variable, the gamma distribution and the associated Laguerre polynomials are chosen. The resulting approximative distribution function[5] $F_q(z) = \mathbb{P}(-x < z)$ by matching the first q moments of $-x$ can be read off from [14, Eq. (2.7.27)] as

$$F_q(z) \approx \frac{\gamma\left(\alpha, z/\beta\right)}{\Gamma(\alpha)} + \epsilon(z),$$

where

$$\epsilon(z) = \sum_{i=3}^{q} w_i \sum_{j=0}^{i}\frac{(-1)^j\Gamma(\alpha+i)}{(i-j)!j!}\frac{\gamma\left(\alpha+j, z/\beta\right)}{\Gamma(\alpha+j)}$$

with

$$w_i = \sum_{l=0}^{i}(-1)^l\frac{\Gamma(i+1)\mathbb{E}\left[(-x)^l\right]}{(i-l)!l!\Gamma(\alpha+l)\beta^l} \tag{15}$$

and $\gamma(a,b) = \int_0^b t^{a-1}\mathrm{e}^{-t}\mathrm{d}t$ denotes the lower incomplete gamma function. The parameters

$$\alpha = \frac{\mathbb{E}^2\left[-x\right]}{\mathbb{E}\left[x^2\right] - \mathbb{E}^2\left[-x\right]}, \quad \beta = \frac{\mathbb{E}\left[x^2\right] - \mathbb{E}^2\left[-x\right]}{\mathbb{E}\left[-x\right]}$$

are obtained by matching the first two moments of $-x$ to a gamma random variable having a density $p(y|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}\mathrm{e}^{-\frac{y}{\beta}}$, $y \in [0,\infty)$. Finally, by the relation (2a), an approximation to the outage probability (3) is obtained as

$$P_{\text{out}}(z) = 1 - F_q\left(m\ln r - z\right). \tag{16}$$

As the number of moments involved in (15) increases, the accuracy of the approximation (16) is expected to improve.

Fig. 1 shows the outage probability of the high-SNR mutual information $\mathcal{I}$ assuming the channel dimensions $m = 4$ and $n = 6$, where different dimensions of the scattering matrices

[3]The corresponding first two moments and some bounds on the outage probability have been derived in [13]

[4]The corresponding first and second moment has been obtained in [7, Eq. (15)] and [9, Eq. (105)], respectively.

[5]Note that the dependence on q is through the summation index in $\epsilon(z)$.

Gharavi, Hamid.
"Exact Moments of Mutual Information of Jacobi MIMO Channels in High-SNR Regime."
Paper presented at IEEE Global Communications Conference 2018 (GLOBECOM 2018), Abu Dhabi, United Arab Emirates. December 9, 2018 - December 13, 2018.

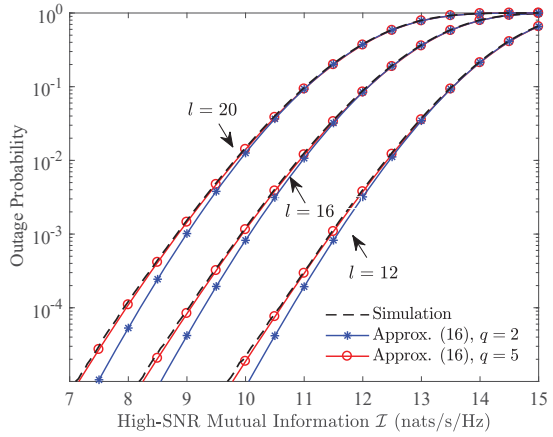


Fig. 1. Outage probability of high-SNR mutual information (2) for different l with $m = 4$, $n = 6$, and $r = 20$ dB.

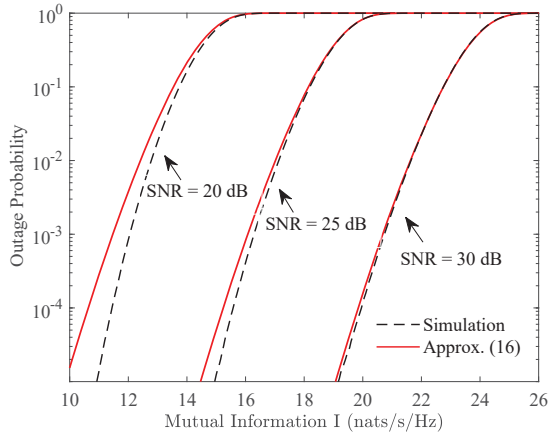


Fig. 2. Outage probability of mutual information (1) for different SNR values with $q = 5$, $l = 12$, $m = 4$, and $n = 6$.

$l = 12$, 16, and 20 are considered. The numerical simulations are compared with the moment-based approximative outage probability (16), where the number of moments considered are $q = 2$ and $q = 5$. We see that the outage probability decreases as the number of untapped channels $l - m - n$ decreases. This phenomenon is also observed in [6]. As expected, it is seen that the accuracy of the proposed approximation (16) improves especially in the tails as the number of moments used increases from $q = 2$ to $q = 5$.

The impact of finite SNR on the accuracy of the approximation (16) is evaluated in Fig. 2, where the number of moments used is $q = 5$ and the channel dimensions are $l = 12$, $m = 4$, and $n = 6$. Despite the asymptotic nature of the approximation, we see that it is already reasonably accurate for not-so-high SNR for outage probability as low as 10^{-4}.

V. Conclusion

We study the mutual information of the Jacobi MIMO channels, which is a realistic model for optical and interference-limited multiuser communications. The corresponding exact moments for such a channel model in the high-SNR regime are derived. The results are possible by making use of the matrix integrals involving the density of the high-SNR mutual information. Approximations to the outage probability are constructed based on the obtained exact moments. Simulation demonstrates the accuracy of the results in practical scenarios of low outage probability and finite number of antennas.

References

[1] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.,* vol. 10, no. 6, pp. 585-595, Nov. 1999.

[2] R. Dar, M. Feder, and M. Shtaif, "The Jacobi MIMO channels," *IEEE Trans. Inf. Theory,* vol. 59, no. 4, pp. 2426-2441, Apr. 2013.

[3] G. Akemann, M. Kieburg, and L. Wei, "Singular value correlation functions for products of Wishart random matrices," *J. Phys. A: Math. Theor.,* 46, 275205, 2013.

[4] G. Akemann, J. R. Ipsen, and M. Kieburg, "Products of rectangular random matrices: singular values and progressive scattering," *Phys. Rev. E,* 88, 052118, 2013.

[5] Y. Chen and M. R. McKay, "Coulumb fluid, Painlevé transcendents, and the information theory of MIMO systems," *IEEE Trans. Inf. Theory,* vol. 58, no. 7, pp. 4594-4634, July 2012.

[6] A. Karadimitrakis, A. L. Moustakas, and P. Vivo, "Outage capacity for the optical MIMO channel," *IEEE Trans. Inf. Theory,* vol. 60, no. 7, pp. 4370-4382, July 2014.

[7] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inf. Theory,* vol. 51, no. 12, pp. 4134-4151, Dec. 2005.

[8] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Arguments.* Singapore: World Scientific, 1997.

[9] M. R. McKay and I. B. Collings, "General capacity bounds for spatially correlated Rician MIMO channels," *IEEE Trans. Inf. Theory,* vol. 51, no. 9, pp. 3121-3145, Sept. 2005.

[10] G. E. Andrews, *The Theory of Partitions.* New York: Cambridge University Press, 1984.

[11] Y. L. Luke, *The Special Functions and Their Approximations. Vol. I.* New York: Academic Press, 1969.

[12] L. Wei, "Proof of Vivo-Pato-Oshanin's conjecture on the fluctuation of von Neumann entropy," *Phys. Rev. E,* 96, 022106, 2017.

[13] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.,* vol. 51, no. 11, pp. 2784-2795, Nov. 2003.

[14] H. T. Ha, *Advances in Moment-Based Density Approximation Methods.* Ph.D. thesis, University of Western Ontario, 2006.

[15] S. Provost and H. T. Ha, "On the inversion of certain moment matrices," *Linear Algebra Appl.,* vol. 430, no. 10, pp. 2650-2658, May 2009.

Appendix

Before proving Proposition 1, we need the following lemma on the finite sums of polygamma functions.

Lemma 1. *For a positive integer l, we have*

$$\sum_{k=1}^{n} \psi_0(k+l) = (n+l)\psi_0(n+l) - l\psi_0(l) - n, \quad (17a)$$

$$\sum_{k=1}^{n} \psi_i(k+l) = (n+l)\psi_i(n+l) - l\psi_i(l) + i\left(\psi_{i-1}(n+l) - \psi_{i-1}(l)\right), \quad i \geq 1. \quad (17b)$$

Gharavi, Hamid.
"Exact Moments of Mutual Information of Jacobi MIMO Channels in High-SNR Regime."
Paper presented at IEEE Global Communications Conference 2018 (GLOBECOM 2018), Abu Dhabi, United Arab Emirates. December 9, 2018 - December 13, 2018.

Proof: By the definition of digamma function (14a), we have

$$\psi_0(k+l) = \psi_0(l) + \sum_{i=0}^{k-1} \frac{1}{l+i}, \tag{18}$$

which gives

$$\begin{aligned}
&\sum_{k=1}^{n} \psi_0(k+l) \\
&= n\psi_0(l) + \sum_{k=1}^{n}\sum_{i=0}^{k-1} \frac{1}{l+i} \\
&= n\psi_0(l) + \sum_{i=0}^{n-1}\sum_{k=i+1}^{n} \frac{1}{l+i} \\
&= n\psi_0(l) + n\sum_{i=0}^{n-1} \frac{1}{l+i} - \sum_{i=1}^{n-1} \frac{i}{l+i} \\
&= n\psi_0(l) + n\left(\psi_0(n+l) - \psi_0(l)\right) - \sum_{i=1}^{n-1} \frac{l+i-l}{l+i} \\
&= n\psi_0(n+l) + l\sum_{i=1}^{n-1} \frac{1}{l+i} - n + 1 \\
&= n\psi_0(n+l) + l\left(\psi_0(n+l) - \psi_0(l+1)\right) - n + 1 \\
&= (n+l)\psi_0(n+l) - l\psi_0(l) - n.
\end{aligned}$$

This proves (17a). To show (17b), by the series representation of polygamma function (13), one has

$$\psi_i(k+l) = \psi_i(l) + (-1)^i i! \sum_{j=0}^{k-1} \frac{1}{(l+j)^{i+1}}, \tag{19}$$

which similarly gives

$$\begin{aligned}
&\sum_{k=1}^{n} \psi_i(k+l) = n\psi_i(l) + (-1)^i i! \sum_{j=0}^{n-1}\sum_{k=j+1}^{n} \frac{1}{(l+j)^{i+1}} \\
&= n\psi_i(l) + (-1)^i i! n \sum_{j=0}^{n-1} \frac{1}{(l+j)^{i+1}} - (-1)^i i! \sum_{j=1}^{n-1} \frac{j}{(l+j)^{i+1}} \\
&= n\psi_i(l) + n\left(\psi_i(n+l) - \psi_i(l)\right) - (-1)^i i! \sum_{j=1}^{n-1} \frac{l+j-l}{(l+j)^{i+1}} \\
&= n\psi_i(n+l) - (-1)^i i! \sum_{j=1}^{n-1} \frac{1}{(l+j)^{i}} + (-1)^i i! l \sum_{j=1}^{n-1} \frac{1}{(l+j)^{i+1}} \\
&= n\psi_i(n+l) + i\left(\psi_{i-1}(n+l) - \psi_{i-1}(l) - \frac{(-1)^{i-1}(i-1)!}{l^i}\right) \\
&\quad + l\left(\psi_i(n+l) - \psi_i(l) - \frac{(-1)^i i!}{l^{i+1}}\right) \\
&= (n+l)\psi_i(n+l) - l\psi_i(l) + i\left(\psi_{i-1}(n+l) - \psi_{i-1}(l)\right).
\end{aligned}$$

■

Note that the equality (17a) and the special case $i = 1$ of the equality (17b) appeared in [12, Eq. (A1)] and [12, Eq. (A7)], respectively, where no proofs were provided.

With the results in Lemma 1, we now turn to the proof of Proposition 1.

Proof: As previously mentioned, the key ingredient of our results is to make use of the relevant matrix integral instead of the integral over the eigenvalue density (5). Mathematically, the Jacobi MIMO channel (4) is a specific truncation of an unitary matrix, where the corresponding density of the channel matrix $\mathbf{H}\mathbf{H}^\dagger$ is given by [8, p. 357]

$$\frac{\Gamma_m(\alpha+l-n)}{\Gamma_m(\alpha)\Gamma_m(l-n)} \det\left(\mathbf{H}\mathbf{H}^\dagger\right)^{\alpha-m} \det\left(\mathbf{I}_m - \mathbf{H}\mathbf{H}^\dagger\right)^{l-m-n}. \tag{20}$$

Here, $\Gamma_m(\alpha)$ denotes the multivariate gamma function [8]

$$\Gamma_m(\alpha) = \pi^{\frac{1}{2}m(m-1)} \prod_{k=0}^{m-1} \Gamma(\alpha-k) \tag{21}$$

and the parameter α in the density (20) equals n.

Now the cumulant generating function (7) of the random variable (6) over the matrix density (20) is calculated as

$$\begin{aligned}
K(s) &= \ln \mathbb{E}\left[\mathrm{e}^{s \ln \det\left(\mathbf{H}\mathbf{H}^\dagger\right)}\right] \\
&= \ln \mathbb{E}\left[\det\left(\mathbf{H}\mathbf{H}^\dagger\right)^s\right] \\
&= \ln \frac{\Gamma_m(\alpha+l-n)}{\Gamma_m(\alpha)\Gamma_m(l-n)} + \ln \int \det\left(\mathbf{H}\mathbf{H}^\dagger\right)^{s+\alpha-m} \times \\
&\quad \det\left(\mathbf{I}_m - \mathbf{H}\mathbf{H}^\dagger\right)^{l-m-n} \mathrm{d}\mathbf{H}\mathbf{H}^\dagger \\
&= \ln \frac{\Gamma_m(\alpha+l-n)}{\Gamma_m(\alpha)\Gamma_m(l-n)} + \ln \frac{\Gamma_m(s+\alpha)\Gamma_m(l-n)}{\Gamma_m(s+\alpha+l-n)}.
\end{aligned}$$

The integral in the above can be considered as a trivial deformation of the density (20), which is directly obtained by replacing in the normalization constant the appearance of α by $s+\alpha$. Setting $\alpha = n$, according to the definition (8) the i-th cumulant $\tilde{\kappa}_i$ of (6) can now be computed as

$$\begin{aligned}
\tilde{\kappa}_i &= \frac{\mathrm{d}^i}{\mathrm{d}s^i} K(s)\Big|_{s=0} \\
&= \frac{\mathrm{d}^i}{\mathrm{d}s^i} \ln \frac{\Gamma_m(s+n)}{\Gamma_m(s+l)}\Big|_{s=0} \\
&= \frac{\mathrm{d}^i}{\mathrm{d}s^i} \left(\sum_{k=0}^{m-1} \ln\Gamma(s+n-k) - \sum_{k=0}^{m-1} \ln\Gamma(s+l-k)\right)\Big|_{s=0} \\
&= \sum_{k=0}^{m-1} \psi_{i-1}(n-k) - \sum_{k=0}^{m-1} \psi_{i-1}(l-k) \\
&= \sum_{k=1}^{m} \psi_{i-1}(n-m+k) - \sum_{k=1}^{m} \psi_{i-1}(l-m+k),
\end{aligned}$$

where we have used the definitions (21) and (13). Finally, by using Lemma 1 and the relation between the cumulants (9), we prove Proposition 1. ■

It is seen that instead of directly studying the moments, the use of cumulant generating function turns out to be more convenient, where all the cumulants are obtained at once via the polygamma functions.

Gharavi, Hamid.
"Exact Moments of Mutual Information of Jacobi MIMO Channels in High-SNR Regime."
Paper presented at IEEE Global Communications Conference 2018 (GLOBECOM 2018), Abu Dhabi, United Arab Emirates. December 9, 2018 - December 13, 2018.

Building a Beacon Format Standard: An Exercise in Limiting the Power of a TTP

No Author Given

No Institute Given

Abstract. We discuss the development of a new format for beacons–servers which provide a sequence of digitally signed and hash-chained public random numbers on a fixed schedule. Users of beacons rely on the trustworthiness of the beacon operators. We consider several possible attacks on the users by the beacon operators, and discuss defenses against those attacks that have been incorporated into the new beacon format. We then analyze and quantify the effectiveness of those defenses.

1 Introduction

A randomness beacon is a source of timestamped, signed random numbers; randomness beacons (among other kinds) were first described by Rabin in [7].

At high level, a randomness beacon is a service that regularly outputs randomness, with certain cryptographic guarantees and with associated metadata, including a timestamp and a cryptographic signature. Each output of a beacon is called a *pulse*, and the sequence of all associated pulses is called a *chain*. An individual pulse from a given beacon can be uniquely identified by its `timeStamp`, and a pulse must never become visible before its timestamp.

In 2013, NIST set up a prototype beacon[5] service. Recently, a new beacon format has been developed with a number of new security features. Several independent organizations are currently planning to adopt this format, which offers the opportunity to have multiple independent beacon operators (in different countries) supporting an identical format, and supporting the generation of random values using multiple beacons inputs.

A beacon is a kind of *trusted third party*(TTP)–an entity which is trusted to behave properly, in order to get some cryptographic protocol to work securely. It's very commonly the case that we need a TTP to get a cryptographic protocol to work, or to make it efficient. However, this introduces a new security concern: the TTP is usually in a position to violate the security guarantees of the system. In the case of a beacon, users of a beacon gain the benefits of a convenient source of public random numbers, but must now worry that the beacon operator may reveal those random numbers in advance to favored parties, manipulate the public random numbers for his own purposes, or even "rewrite history" by lying about the value of previous beacon pulses.

Kelsey, John.
"Building a Beacon Format Standard: An Exercise in Limiting the Power of a Trusted Third Party (TTP)."
Paper presented at SSR 2018: Security Standardisation Research, Darmstadt, Germany. November 26, 2018 - November 27, 2018.

The design of the new beacon format was largely an exercise in trying to limit the power of a TTP (the beacon) to violate its users' security. We believe this effort has some lessons for anyone trying to design a cryptographic protocol or standard that requires a trusted third party.
In the remainder of this document, we first propose a few principles for designing protocols with trusted third parties to minimize their potential harm. We then discuss the current NIST beacon and the general applications of public randomness. Next, we illustrate our TTP-limiting principles with specific aspects of the design of the NIST beacon format. We conclude by considering the ultimate effect of our efforts, and consider some possible improvements for the future.

2 Trusted Third Parties

A trusted third party (TTP) is an entity in some cryptographic protocol who must behave properly, in order for the protocol to meet its security goals. For example, in a public key infrastructure, a certificate authority (CA) is a TTP. The CA's job is to verify the identity of users, and then to issue users a certificate that binds their identity to a public key. A CA which is careless or malevolent can issue certificates to the wrong users–for example, giving an attacker a certificate that lets him impersonate an honest user. In most currently-used voting systems, the voting hardware is a kind of trusted third party; if it misbehaves, it may undermine the security of the election. A good discussion of the downsides of TTPs appears in [12].
Many important real-world cryptographic systems require TTPs to function. However, a TTP in a system is always a source of vulnerabilities to attack–if the TTP misbehaves, the security of the system will be violated. Because of this, anyone designing or standardizing a cryptographic system with a TTP should try to minimize the potential for abuse or misbehavior of the TTP. In general, this requires going through a few steps:

1. Enumerating each way that the TTP might misbehave so as to undermine the security claims of the system.
2. For each such potential attack, adding features to mitigate the attacks as much as possible.
   (a) In some cases, the attack may be possible to eliminate entirely. This essentially means removing one area of required trust from the TTP.
   (b) Sometimes, it may not be possible to eliminate the attack, but it may still be possible to limit its power or scope.
   (c) In other cases, the design of the TTP or protocol can be altered so that misbehavior creates evidence that can be shown to the world.
   (d) In still other cases, the design may be altered so that misbehavior is at least likely to be *detected* when it is attempted, even if this produces no solid evidence that can be shown to others.
   (e) Sometimes, an external or optional auditing step can be added that makes the misbehavior likely to be detected.

Kelsey, John.
"Building a Beacon Format Standard: An Exercise in Limiting the Power of a Trusted Third Party (TTP)."
Paper presented at SSR 2018: Security Standardisation Research, Darmstadt, Germany. November 26, 2018 - November 27, 2018.

(f) In the worst case, it may be impossible to make this kind of misbehavior impossible or even detectable–in that case, we can at least document the risk of this kind of misbehavior.

Each of these steps is valuable. Once the designer of a system realizes that the TTP could misbehave in some way, it is often possible to work out mechanisms to move up the list–to make an undetectable attack detectable, or to make some kind of misbehavior by the TTP leave evidence that can be shown to the world. Even the final step, documenting the risk of misbehavior, is worthwhile, as it allows users of the TTP to understand exactly what risks they are accepting.

When there is some auditing step added to the protocol to keep the TTP honest, it is very important to consider how practical the auditing step is. If it is extremely computationally expensive or otherwise inconvenient, it may never be done even if it is, in principle, possible.

2.1 Incentives for the Honest TTP

An entity that wants to run an honest TTP has many strong incentives to try to reduce its own scope for abuse.

Fear, Uncertainty, and Doubt When silent, undetectable misbehavior is possible, there will always be suspicions swirling around that it is being done. This can undermine the TTP's reputation, or even convince people to avoid the system that relies on the TTP. Sometimes, this will happen even when the alternative is obviously less secure.

Insider and Outsider Attacks An organization trying to run a TTP honestly must deal with the possibility of both insider and outsider attacks. An insider might cause the TTP to misbehave in some way, despite the good intentions of the organization running the TTP. Or an outsider might subvert the security of the TTP in some way. Both of these have the potential to damage the reputation of the organization running the TTP.

Coercion The organization running a TTP must also worry about the possibility that they will be coerced in some way to misbehave. This might be done via legal means (such as a national security letter), or extralegal ones (a mobster threatening the head of the organization if some misbehavior isn't done). It might involve financial pressure (a business partner threatening some dire reprisals if the misbehavior isn't done), or a change in management of the organization in the future. It could even involve some intense short-term temptation to do the wrong thing that might be hard to resist in the future. The best way to resist this coercion is to have bad behavior simply be unworkable; what you cannot do, you cannot be coerced to do.

This follows the general pattern noted in [9], in which it is sometimes possible to improve one's position by limiting one's options. When it comes to the design and operation of a TTP, both the users and the operators of a TTP benefit from having a TTP and surrounding system designed to minimize the scope for abuse, and to make any abuse instantly detectable.

Kelsey, John.
"Building a Beacon Format Standard: An Exercise in Limiting the Power of a Trusted Third Party (TTP)."
Paper presented at SSR 2018: Security Standardisation Research, Darmstadt, Germany. November 26, 2018 - November 27, 2018.

3 Public Randomness and the NIST Beacon

As described above, a beacon is a source of *public random numbers*, released in self-contained messages called *pulses*. The numbers should be random–unpredictable and impossible for anyone to influence. Each pulse contains a timestamp, and neither the pulse nor its random values may be released before the time indicated in the timestamp. Pulses are digitally signed and incorporated into a hash chain to guarantee their integrity. In order to be useful, a beacon should issue pulses on a fixed schedule, and should keep a database of all old pulses which it will provide on request.

3.1 Public Random Numbers?

Most discussions of random numbers in cryptography involve secret random numbers, such as those used to generate keys. However, there are many places where it is more useful to have *public* random numbers–numbers which are not known by anyone until some specific time, and then later become known (or at least available) to the whole world.
Why would we want *public* random numbers?

- Demonstrate we did something randomly (outside our own control).
- Bind something in time. (It couldn't have happened before time T.)
- Coordinate a random choice among many parties.
- Save messages in a crypto protocol.

The added value of a beacon is that these random numbers can be shown even to parties who were not present or involved at the time some event occurred. For example, if the random numbers from a beacon pulse are used to select a subset of shipping containers to open for inspection, then people who weren't present and weren't participating can verify that the selection was random–assuming they trust the beacon.
There are a number of properties that users need from public random numbers:

- Genuinely random.
- Unpredictable to anyone until they become public.
- Verifiable by everyone after they become public.
- Impossible for anyone to control or manipulate.
- Impossible to alter past values.

3.2 A Beacon as a TTP

All these applications work well with a beacon[1]. However, users of the beacon (including anyone trusting the randomness of the choices made) need to trust that the beacon is behaving correctly. A corrupt beacon might do all kinds of bad things, such as:

[1] There are other ways to get public random numbers. Many also depend on some trusted third party; others introduce other practical problems–ambiguity about correct values, lack of a fixed schedule, etc. Overall, we believe beacons are the best way to get practical public randomness for real-world applications

Kelsey, John.
"Building a Beacon Format Standard: An Exercise in Limiting the Power of a Trusted Third Party (TTP)."
Paper presented at SSR 2018: Security Standardisation Research, Darmstadt, Germany. November 26, 2018 - November 27, 2018.

– Reveal future random values early to favored people.
– Control or influence random values to undermine the security of some user.
– "Rewrite history" by altering old pulses, thus changing claimed historical results.

The NIST beacon is a practical source of public random numbers that has been running continuously (with occasional downtime) for five years. The NIST beacon consists of several parts:

– Engine: the device that actually generates the pulses.
– Multiple independent cryptographic RNGs used to generate random values.
– A commercial hardware security module (HSM) used both as one source of random bits, and also to sign pulses.
– Frontend–the machine talking to the world, providing the random numbers. The frontend must support various requests from users.

The beacon format used since 2013 had only a few fields:

– Version
– Frequency
– Seed Value
– Previous Output
– Signature
– Status
– OutputValue

The new format has added many new fields. In figure 1, the new fields appear in red.

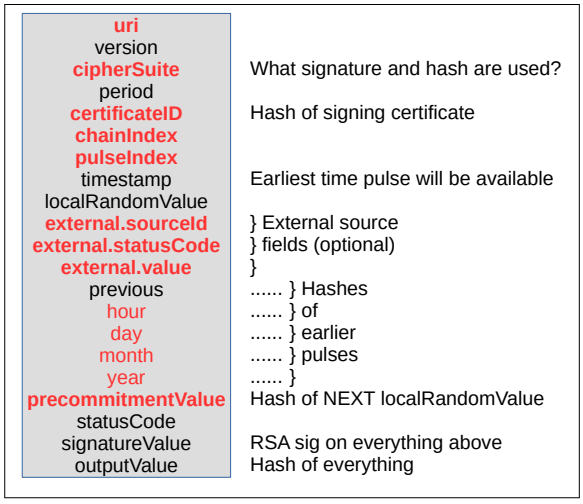


Fig. 1. The new pulse format

Despite the complexity of the new format, there are only a few fields in each pulse which are actually important for users:

– `timeStamp` – the timestamp (UTC) at which the value will be released.
– `localRandomValue` – the internal random value produced once a minute.
– `signatureValue` – a digital signature on the pulse contents.
– `outputValue` – the output of the pulse–the hash of all other fields of the pulse.

The other fields (including all the new fields) add security or improve convenience in using the pulses.

3.3 Timestamps

The `timeStamp` in the new pulse format is always in UTC, so that the time at which a pulse is created is easy for any user to determine, without the need to determine time zone or daylight savings time rules. The beacon promises never to release a pulse (or any information about the `localRandomValue` or `outputValue` field from the pulse) before the time in the timestamp, and never to issue more than one pulse with a given timestamp. The beacon also issues pulses on a fixed timetable–each timestamp is separated by `period` milliseconds. In the new format, `timeStamp` always appears as a string in RFC3339[**RFC3336**] format.

3.4 Where the Random Numbers Come From

Each pulse contains a new random value, called `localRandomValue`. This value is generated using multiple independent, strong random number sources. At present, the NIST beacon uses two RNGs–the Intel RNG included in the Beacon Engine machine, and the hardware RNG inside the HSM. We are currently working on adding a third RNG based on quantum phenomena (single-photon detection), custom built by NIST scientists.

Suppose the beacon has k different RNGs. Let R_i be the 512-bit random output from the i-th RNG. Then, we have the following formula:

$$\texttt{localRandomValue} \leftarrow \texttt{SHA512}(R_0 \parallel R_1 \parallel \ldots \parallel R_{k-1}) \tag{1}$$

That is, we take 512 bits from each independent RNG, concatenate them, and then hash them with SHA512[10]. The resulting 512-bit value becomes `localRandomValue` in the pulse. If *any* of the internal RNGs generate unpredictable, random values, then the result will also be unpredictable and random. Even if one of the RNGs is flawed or backdoored, the value of `localRandomValue` will still be secure, as long as at least one of the RNGs is good.

However, there is no way for any outside observer to verify `localRandomValue` is generated in this way. If the beacon began simply putting any value it wanted in that field, there would be no way for users to detect this.

3.5 Signing Pulses

At any given time, the beacon is using a particular signing key. The beacon also has a corresponding certificate, issued by a well-known commercial CA. (The beacon frontend has an entirely different certificate and key used so that it can support TLS; there is no relationship between these keys.)

In the new format, each pulse contains:

- `signatureValue` is an RSA[11] signature over the entire contents of the pulse.
- `certificateId` contains the `SHA512`() of the X.509[3] certificate of the signing key used for pulses. (The full X.509 certificate is available from the beacon frontend.)

Having each beacon pulse digitally signed accomplishes two security goals: First, it ensures that impersonating the beacon will be difficult. Second, it ensures that a beacon which issues invalid pulses is creating permanent evidence of its misbehavior.

In the NIST beacon, the RSA key used for signing the pulse is kept inside a commercial HSM. The HSM is designed to prevent the beacon engine from extracting the RSA key. This limits the ability of some attackers (both insiders and outsiders) to carry out some attacks, as discussed in detail below.

3.6 The Output Value

The current beacon format (like the previous one) defines the output value of the pulse as the final field in the pulse, labeled `outputValue`. The output value is defined as

$$\texttt{outputValue} \leftarrow \texttt{SHA512}(\text{ all other fields in the pulse }) \qquad (2)$$

Because this includes the value of `localRandomValue`, `outputValue` contains the all the randomness in the pulse. However, users of the beacon can always verify that `outputValue` was computed correctly. As described below, the fact that users of the beacon will normally use `outputValue` has important security consequences.

4 Limiting the Beacon's Power

4.1 Attacks

Above, we noted the sorts of bad behavior possible for a corrupt beacon operator:

- Reveal future random values early to favored people.
- Control or influence random values to undermine the security of some user.
- "Rewrite history" by altering old pulses, thus changing claimed historical results.

Kelsey, John.
"Building a Beacon Format Standard: An Exercise in Limiting the Power of a Trusted Third Party (TTP)."
Paper presented at SSR 2018: Security Standardisation Research, Darmstadt, Germany. November 26, 2018 - November 27, 2018.

A good design of the beacon format and the normal operations of a beacon should ideally make such violations impossible; if not, it should at least make them harder, or at least detectable. In this section, we discuss mechanisms in the new beacon format which either add difficulty to misbehavior by the beacon, or which support users of the pulses protecting themselves from misbehavior.

4.2 Prediction: Revealing the Value of `outputValue` in the Future

The first way a beacon operator can misbehave is to reveal future outputs to selected people.

A beacon that is operating correctly generates a new random value very soon before issuing a given pulse, and so can't possibly tell anyone a value of the pulse very far into the future. However, a misbehaving beacon operator can, in principle, generate its pulses far in advance. All the computations needed to compute a pulse are relatively cheap, so computing a year's worth of pulses in advance can easily be done in an hour or two.

There are two limits on the ability of a corrupt beacon to reveal future pulses:

The Certificate ID The `certificateId` is the `SHA512`() of the certificate currently being used to sign pulses. Certificates normally have a fixed validity period–for example, one year. A corrupt beacon operator cannot predict the value of a certificate until he has seen it. Therefore, the validity period of the certificate imposes a (very generous) limit on the ability of the beacon to precompute future pulses.

Suppose each certificate has a validity period of one year, and is issued one month ahead of its validity period. Then the beacon is limited to precomputing pulses no more than 13 months in advance.

The External Source The new format allows beacons to optionally incorporate an *external source* into pulses. For example, a given beacon could incorporate the winning Powerball lottery numbers twice per week, or the closing value of the DJIA at the end of every stock trading day. By incorporating an external source that is outside the control of the beacon, the beacon operator can demonstrably lose a huge amount of power.

Each pulse contains a field called `external.value`, which contains the `SHA512`() of the current external value. It is updated only when a new external value appears; otherwise, it retains the value it had previously.

Each pulse also contains a field called `external.sourceId`, which contains the `SHA512`() of a text description of the external source and how and when the external value will be updated. This should almost never change.

Suppose a beacon incorporates a lottery drawing which occurs once per day as its external source; the result is incorporated six hours after the result is made public. In this case, the beacon is limited to precomputing future values no more than 30 hours in advance.

Prediction: Summary

1. A corrupt beacon can precompute its pulses far in advance and reveal the output values to friends.
2. Incorporating an optional `external.value` into the pulse limits the precomputation (and thus the ability to reveal future outputs) to the time between updating the `external.value`.

4.3 Influencing Outputs

Recall that the output of the pulse is computed as:

$$\texttt{outputValue} \leftarrow \texttt{SHA512}(\text{all other fields of the pulse})$$

Further, the recipient of the pulse can verify that this value is correct by recomputing the hash. This means that, while the beacon operator can completely control the value of `localRandomValue`, he cannot directly control the value of `outputValue`. Instead, in order to exert control over this value, he must try many different inputs; each input leads to a different random output value. With 2^k tries, he can expect to get any property he likes for `outputValue` which has a probability no lower than 2^{-k}.

For concreteness, in the rest of this discussion, we will assume that the beacon operator wants to force the least significant bits of a given pulse's `outputValue` to be zeros. Thus, with 2^k work, he can expect to force the low k bits to zero. This generalizes; an output value with any property that has a 2^{-k} probability can be found about as easily.

From [1], we can get benchmarks for an 8-GPU desktop machine doing brute-force hashing[2]. The listed system can do about 8.6 million SHA512 hashes per second, which is about 2^{23} hashes per second. An attacker with such a system could do about 2^{39} hashes per day, or 2^{47} hashes per year. Because the whole process is parallelizeable, spending N times as much money will get an N-way speedup. However, in order to add one bit of control of `outputValue` with the same hardware, an attacker must double the time taken for the computation.

At the time of this writing, the most powerful attacker imaginable might conceivably have the computing resources to compute 2^{90} SHA512 hashes in a year. This puts an upper bound on the beacon operator's control of `outputValue`. However, for concreteness, we will assume in the rest of this discussion that the attacker has the 8-GPU desktop machine described in [1], and is trying to control bits of `outputValue` using it.

There are several components of the new pulse format which limit the attacker's control of the output bits.

The Certificate ID As described above, the beacon operator can't predict the value of its next certificate, incorporated into every pulse in `certificateId`. This limits the attacker to no more than about 13 months of precomputation–with only about a year to do his attack, he is limited to controlling 47 bits of `outputValue`.

[2] That system is doing password cracking attacks. Trying to control some bits of the output of the beacon is very similar to password cracking.

The Signature Each pulse contains an RSA signature, `signatureValue`, over all fields except the `outputValue`. This affects the attack in two ways: First, it means that each new value of a pulse will require a new RSA signature–adding slightly to the work needed per new output value computed. This probably reduces the attacker's ability to control `outputValue` by a small number of bits.

More importantly, in the NIST beacon, the RSA signing key is stored inside an HSM, and should be very difficult to extract. If the attacker doesn't have access to the signing key[3], then the attacker must involve the HSM in every new computation of an output value.

Suppose the HSM can do 100 RSA signatures per second. Then, the attacker is limited to only about 2^{32} trial values for the output in a year of trying, and so can control only about 32 bits of `outputValue`.

`external.value` As described above, the new format allows beacons to optionally incorporate an *external source* into pulses.

Once again, suppose the external source is updated from a lottery drawing carried out each day. The lottery drawing results become public six hours before the value is incorporated into the `external.value` field. (This must be described precisely in the text whose hash appears in `external.sourceId`.) The attacker's ability to control outputs is now enormously diminished–his precomputations can now run for at most 30 hours.

An attacker who has extracted the RSA key from the HSM and has the 8-GPU system can thus carry out no more than 2^{40} computations, and so can control about 40 bits of `outputValue`. An attacker who hasn't extracted the RSA key from the HSM can control only about 23 bits of the output value.

This makes a pretty good argument for the idea that the private key for the beacon should be generated inside the HSM once and never released to anyone, even encrypted. In the best case, we'd have some kind of guarantee that even the beacon operator could never see the private key.

Influence: Summary Consider an attacker trying to control `outputValue`:

1. A vastly powerful and well-funded attacker might be able to control as many as 90 bits given a year of computation.
2. An attacker with a powerful setup for password cracking can control 47 bits in a year of computation.
3. When the attacker is unable to extract the RSA key from the HSM, his attack is limited by the speed of RSA signatures from the HSM. If the HSM can do 100/second, then the attacker can control about 40 bits of the output value with a year of computation.
4. If the beacon incorporates external values, the attacker's power is massively diminished. Assuming an external value known to the attacker for only 30 hours:

[3] This might be an outsider who has compromised the beacon engine, or an insider with access to the engine but not the HSM or RSA keys.

(a) With the RSA key from the HSM, the attacker can control 40 bits.
(b) Without the RSA key from the HSM, the attacker can control 23 bits.

It is important to note that there is no way for any user to detect these attacks from the outside–as far as the user is concerned, the beacon is behaving normally, and the pulses look just like any other pulses.

4.4 Combining Beacons

Having multiple beacons operating at the same time gives users a choice of which beacon to trust. However, multiple beacons allow a user to *combine* outputs from two or more beacons. The goal of combining beacons is to reduce trust in the beacon operators: if we combine pulses from N beacons and at least one is honest, then the resulting random value should be random, unpredictable, and outside anyone's control. Instead of having to trust a single beacon operator, the user can end up needing to trust that any one of two or three operators is honest.

Notation for talking about beacons In order to discuss combining of beacons, we need to introduce some notation: Suppose A and B are beacons:

- $A[T]$ is beacon pulse at time T from A.
- $B[T]$ is beacon pulse at time T from B.
- $B[T+1]$ = next pulse, $B[T-1]$ = previous pulse.
- $B[T]$.`localRandomValue` is the `localRandomValue` field of the pulse at time T from beacon B.

Combining Beacons: What doesn't *work?* A natural approach would be to XOR the `outputValue` fields from two beacons, getting

$$Z \leftarrow A[T].\texttt{outputValue} \oplus B[T].\texttt{outputValue}$$

For simplicity, let's suppose A is a corrupt beacon and B is an honest one.

Combining the beacons even in this simple way prevents the prediction attack: A can't know what random number B will produce at time T, so he cannot reveal the future Z to his friends.

Unfortunately, A can still exert control over the combined output Z. A uses the following strategy to exert control over the bits of Z, despite B's genuinely random contribution:

1. B does a precomputation of 2^{32} possible values of $B[T]$.`outputValue`.
2. A sends pulse $A[T]$.
3. B observes $A[T]$.`outputValue`
4. B chooses $B[T]$ to control low 32 bits of

$$A[T].\texttt{outputValue} \oplus B[T].\texttt{outputValue}$$

.

This attack led us to add a new field to the beacon format: `precommitmentValue`.

$$A[T-1].\texttt{precommitmentValue} = \texttt{SHA512}(A[T].\texttt{localRandomValue})$$

That is, $A[T-1]$ **commits** to `localRandomValue` in $A[T]$. Once a user has seen $A[T-1]$, A has **no choice** about the value of $A[T]$.`localRandomValue`. Note that `precommitmentValue` requires that the beacon engine compute the value of `localRandomValue` one pulse in advance.
By forcing the beacon to commit up front to the next pulse' random value, we can construct a protocol for combining beacons that is much more secure.

Combining Beacons the Right Way In order to combine beacons from A and B, the user does the following steps:

1. Receive $A[T-1]$ and $B[T-1]$.
2. Verify that both pulses:
   (a) Are valid
   (b) Were received before time T
3. Receive $A[T]$ and $B[T]$.
4. Verify that the `precommitmentValue` fields are in agreement with the `localRandomValue` fields:

$$A[T].\texttt{precommitmentValue} = \texttt{SHA512}(A[T].\texttt{localRandomValue})$$
$$B[T].\texttt{precommitmentValue} = \texttt{SHA512}(B[T].\texttt{localRandomValue})$$

5. Compute combined value by hashing together:
   - `outputValue` fields from $T-1$
   - `localRandomValue` fields from T

$$Z \leftarrow \texttt{SHA512}(A[T-1].\texttt{outputValue} \parallel B[T-1].\texttt{outputValue} \parallel A[T].\texttt{localRandomValue} \parallel B[T].\texttt{localRandomValue})$$

This is a somewhat more complex way of combining beacons, but the added complexity adds security:

1. By incorporating $A[T-1].\texttt{outputValue}$ and $B[T-1].\texttt{outputValue}$, we incorporate the external value, RSA signatures, and other security mechanisms described above.
2. By incorporating $A[T].\texttt{localRandomValue}$ and $B[T].\texttt{localRandomValue}$, we incorporate a fresh random number from each beacon. Any honest beacon will provide a completely unpredictable random value.
3. Thanks to the `precommitmentValue` fields in $A[T-1]$ and $B[T-1]$, neither beacon can change the value of their `localRandomValue` fields after they see the random value from the other beacon.

As long as the beacons output their values on time, the result is that if at least one beacon is honest, Z is random, it can't be predicted by anyone before time T, and it cannot be influenced by either beacon.

Hitting the RESET button Combining beacons massively improves security, but isn't quite perfect–a corrupt beacon still has a very small amount of influence.

Once again, suppose A is an evil beacon, and B is an honest one. The attack works as follows:

1. Both beacons send out pulses for time $T-1$.
2. At T, B sends out $A[T]$.
3. A computes combined output Z.
4. If it doesn't like Z, it simulates a failure.

Note that this attack follows a common pattern: simulating a failure that really could happen for innocent reasons. If the corrupt beacon is willing to do this once for a given attempt to combine beacons, then it gets a very small amount of influence on Z–it gets one opportunity to veto a value of Z it finds unacceptable. Along with massively decreasing the power of a corrupt beacon, this attack also forces a corrupt beacon trying to influence the combined output to take a visible action–it must simulate a failure and stop producing pulses for awhile. This is very visible to the user trying to combine beacon pulses, and also leaves a permanent record in its sequence of pulses.

Summary: Combining Beacons

1. Combining beacons massively decreases the scope of a corrupt beacon–as long as at least one beacon used is honest, the user's security is massively improved.
2. Prediction attacks become impossible–a corrupt beacon cannot reveal the value of a future combined value to a friend even one minute in advance.
3. Influence attacks become much harder–a corrupt beacon can veto one combined value, but must do so in a way that is obvious and leaves a permanent record in the chain of pulses.

4.5 Rewriting History

Both the old and new beacon pulse format have two fields that make it impossible for a beacon operator to "rewrite history" without leaving permanent evidence that it has done so.

First, each pulse has a `signatureValue` field, signing the pulse with a key for which the beacon operator has the key and certificate. Second, each pulse contains the `outputValue` of the previous pulse, in its `previous` field. Since the output value is the hash of the contents of the pulse, this means that a sequence of pulses forms a *hash chain.*

A hash chain has the property that introducing a change anywhere in the chain requires changing every block *after the chain.* Figure 2 shows the hash chain (with most fields omitted for clarity) when a change is made to a pulse.

The result of this is that if the beacon operator makes such a change to some pulse in the past, then that pulse's output value becomes inconsistent with the later pulses in the chain–pulses that users have already seen and stored, and that have valid signatures from the beacon operator. A change to any pulse leaves an inconsistency in the chain. Examining the chain will always detect the inconsistency.

4.6 Changing a pulse changes all future pulses!

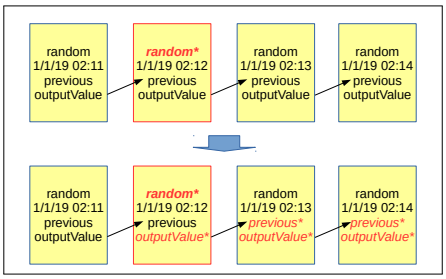


Fig. 2. Altering a pulse affects every future pulse in the hash chain

The result of this is that any attempt to "rewrite history" by changing the value of some past pulse creates permanent evidence of the beacon's misbehavior. Unfortunately, verifying that the beacon hasn't rewritten history can be expensive. Consider the situation where a user know the value of 2022-10-04 17:35, and wants to use this to verify the value of 2019-11-29 22:58. There are about 1.5 million pulses between these two, so a user wanting to do this verification would have to request 1.5 million pulses from the beacon, and then verify the entire hash chain.

Skiplists An auditing step that is unworkably complex and expensive will seldom be done. In order to make this auditing step more efficient, we added several fields to the pulse format, in order to support a much more efficient way to verify that a given pulse has not been changed, by returning a sequence of pulses we refer to as a *skip list*. A skiplist between two pulses, TARGET and ANCHOR, guarantees that pulse TARGET is consistent with the value of pulse ANCHOR–the beacon cannot construct a valid skiplist if the value of TARGET has been changed.
A hash chain used to verify the value of the pulse at 2019-11-29 22:58 given knowledge of the value of the pulse at 2022-10-04 17:35 would require about 1.5 million pulses; the skiplist requires nine pulses. More generally, when the known (ANCHOR) pulse and the pulse to be verified (TARGET) are Y years apart, the skip list will be about $Y + 46.5$ pulses long on average.
The construction of skiplists is described in detail in appendix A.

5 Conclusions and Open Issues

In this article, we have described the design of the new NIST beacon format from the perspective of minimizing the power of a beacon operator to misbehave. By adding fields to the pulse format, and defining a protocol for combining pulses from multiple beacons, we have massively

decreased the power of the beacon operators to misbehave without detection. In the case where a user can combine pulses from two independent beacons, he can be extremely secure from misbehavior by any one beacon operator.

Attack	Resources and Modifications	Vulnerability
Prediction[4]		
	Original format	unlimited
	No external value	13 months
	External value	30 hours
	Combined pulses	attack blocked
Influence[5]		
	Original format	At least 47 bits
	No external, RSA key known	47 bits
	No external, RSA key unknown	32 bits
	External, RSA key known	40 bits
	External, RSA key unknown	23 bits
	Combined pulses	1 bit, but attack is visible
Rewriting History		
	Original format	Permanent evidence, expensive verification
	New format	Permanent evidence, cheap verification

Table 1. What can an evil beacon operator do?

Table 1 shows how various ways for the beacon to operate and be used affects the vulnerability of the users to attacks by a corrupt beacon operator. The important points from this table are:

1. A beacon that includes an external value *enormously* reduces the its own scope for misbehavior.
2. Combining pulses from multiple beacons gives very high security, assuming at least one beacon is honest. (If all the combined beacons are corrupt, then it makes no difference.)

There are two possible directions for further work: we can make changes to future versions of the beacon format, and we can add new recommended protocols for users of the beacon.

5.1 Improving the Beacon Format

The influence attacks described above are computationally expensive. If the output value were computed using a memory-hard function designed to foil password cracking attacks, such as scrypt[6] or Argon 2[2], these attacks would become enormously less practical. Even implementing scrypt with a minimal set of hardness parameters would easily cut ten or more bits off the ability of an attacker to influence outputs.

The "hitting the RESET button" attack exploits the fact that when a beacon commits to its next pulse's `localRandomValue` field, it can still refuse to disclose that. In [4], the authors propose a number of techniques to foil this attack, including the use of time-lock puzzles[8] for the random value in each pulse, and the use of "Merlin chains" to take away any ability for a beacon operator to simulate a failure without permanently shutting down the beacon. We could consider adding these to a future beacon format.

5.2 Improving Protocols and Recommendations for Users

A simple way for a user to protect itself from a corrupt beacon is simply to pre-commit to how it will use the beacon pulse at time T, and also to pre-commit to a secret random number kept by the user. The user reveals his random number at the same time the pulse is revealed, and derives a seed by hashing the pulse's output value and his own random number together. Note that this completely blocks any influence or prediction by the beacon operator, assuming the user is honest.
Instead of adding a memory-hard function to the calculation of the `outputValue`, we could also recommend using a memory-hard function for the computation of the final seed. By choosing the hardness parameters to make this seed calculation take 30 seconds on a reasonably fast desktop machine, the user can massively increase the difficulty of influencing attacks by the beacon operator.

References

[1] *8x Nvidia GTX 1080 Hashcat Benchmarks*. [Online; accessed 09-July-2018].

[2] Alex Biryukov, Daniel Dinu, and Dmitry Khovratovich. "Argon2: New Generation of Memory-Hard Functions for Password Hashing and Other Applications". In: *IEEE European Symposium on Security and Privacy, EuroS&P 2016, Saarbrücken, Germany, March 21-24, 2016*. 2016, pp. 292–302. DOI: 10.1109/EuroSP.2016.31. URL: https://doi.org/10.1109/EuroSP.2016.31.

[3] David Cooper et al. "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile". In: *RFC* 5280 (2008), pp. 1–151. DOI: 10.17487/RFC5280. URL: https://doi.org/10.17487/RFC5280.

[4] Peter Mell, John Kelsey, and James M. Shook. "Cryptocurrency Smart Contracts for Distributed Consensus of Public Randomness". In: *Stabilization, Safety, and Security of Distributed Systems - 19th International Symposium, SSS 2017, Boston, MA, USA, November 5-8, 2017, Proceedings*. 2017, pp. 410–425. DOI: 10.1007/978-3-319-69084-1_31. URL: https://doi.org/10.1007/978-3-319-69084-1%5C_31.

[5] *NIST Randomness Beacon*. https://www.nist.gov/programs-projects/nist-randomness-beacon. [Online; accessed 09-July-2018]. 2018.
[6] Colin Percival and Simon Josefsson. "The scrypt Password-Based Key Derivation Function". In: *RFC* 7914 (2016), pp. 1–16. DOI: 10.17487/RFC7914. URL: https://doi.org/10.17487/RFC7914.
[7] Michael O. Rabin. "Transaction Protection by Beacons". In: *J. Comput. Syst. Sci.* 27.2 (1983), pp. 256–267. DOI: 10.1016/0022-0000(83)90042-9. URL: https://doi.org/10.1016/0022-0000(83)90042-9.
[8] R. L. Rivest, A. Shamir, and D. A. Wagner. *Time-lock Puzzles and Timed-release Crypto*. Tech. rep. Cambridge, MA, USA, 1996.
[9] T. C. Schelling. *The strategy of conflict*. Oxford University Press, 1960.
[10] National Institute of Standards and Technology. *FIPS 180-4, Secure Hash Standard, Federal Information Processing Standard (FIPS), Publication 180-4*. Tech. rep. DEPARTMENT OF COMMERCE, 2015. URL: http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf.
[11] National Institute of Standards and Technology. *FIPS 186-4, Secure Hash Standard, Federal Information Processing Standard (FIPS), Publication 186-4 Digital Signature Standard (DSS*. Tech. rep. DEPARTMENT OF COMMERCE, 2013. URL: http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.186-4.pdf.
[12] Nick Szabo. *Trusted Third Parties are Security Holes*. [Online; accessed 09-July-2018]. 2001.

A Skiplists

In order to support skiplists, we added four additional fields to the pulse format which contain the `outputValue` of previous pulses. (`previous` was defined in the old beacon format.)

`previous` `outputValue` of previous pulse.
`hour` `outputValue` of 1^{st} pulse of *hour* of previous pulse.
`day` `outputValue` of 1^{st} pulse of *day* of previous pulse.
`month` `outputValue` of 1^{st} pulse of *month* of previous pulse.
`year` `outputValue` of 1^{st} pulse of *year* of previous pulse.

By adding these fields, we ensure that instead of a sequence of pulses having a single hash chain, any long sequence of pulses has a huge number of different hash chains running through them, of varying lengths, as shown in figure 3.

The algorithm for extracting a minimal chain (what we call a skiplist) is illustrated in figure 4.

A more precise description of the algorithm appears in algorithm 1.

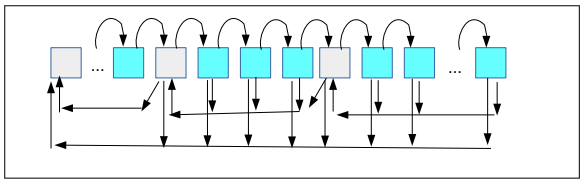


Fig. 3. A long sequence of pulses has a huge number of different hash chains

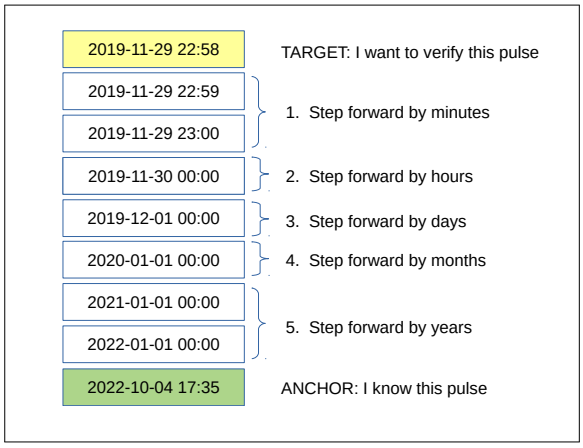


Fig. 4. Constructing a skiplist from TARGET to ANCHOR

In this case, the result is that instead of needing to verify a chain of 1.5 million pulses, the user requests a skiplist of nine pulses from the beacon frontend, and verifies them very efficiently. If the beacon has altered the TARGET pulse, it will not be able to provide a valid skiplist from TARGET to ANCHOR.

In general, a skiplist between a value of TARGET and ANCHOR that are Y years apart will be about $Y + 46.5$ pulses long.

Algorithm 1 Construct a skiplist from TARGET to ANCHOR.

```
1: function MAKE_SKIPLIST(TARGET, ANCHOR)
2:     path ← []
3:     current ← TARGET
4:     while current< ANCHOR do
5:         path ← pp ||current
6:         if current is first pulse in its year then
7:             current ← first pulse in NEXT year
8:         else if current is first pulse in its month then
9:             current ← first pulse in NEXT month
10:        else if current is first pulse in its day then
11:            current ← first pulse in NEXT day
12:        else if current is first pulse in its hour then
13:            current ← first pulse in NEXT hour
14:        else
15:            current ← NEXT pulse
16:    path ← path ||ANCHOR
17:    return (path)
```

Design of superconducting optoelectronic networks for neuromorphic computing

Sonia Buckley, Adam N. McCaughan,
Jeff Chiles, Richard P. Mirin,
Sae Woo Nam, Jeffrey M. Shainline
National Institute of Standards and Technology
Boulder, CO
sonia.buckley@nist.gov

Grant Bruer, James S. Plank
Department of EECS
University of Tennessee
Knoxville, TN
jplank@utk.edu

Catherine D. Schuman
Computational Data Analytics
Oak Ridge National Laboratory
Oak Ridge, TN
schumancd@ornl.gov

***Abstract*—We have previously proposed a novel hardware platform (SOEN) for neuromorphic computing based on superconducting optoelectronics that presents many of the features necessary for information processing in the brain. Here we discuss the design and training of networks of neurons and synapses based on this technology. We present circuit models for the simplest neurons and synapses that we can use to build networks. We discuss the further abstracted integrate and fire model that we use for evolutionary optimization of small networks of these neurons. We show that we can use the TENNLab evolutionary optimization programming framework to design small networks for logic, control and classification tasks. We plan to use the results as feedback to inform our neuron design.**

I. Introduction

The human brain is capable of complex pattern recognition and learning tasks with a fraction of the power consumption of a traditional computer. Neuromorphic computing aims to build hardware that emulates the brain to provide some of this advantage [1]. Deep learning techniques inspired by the brain have already proven to be very effective for many pattern recognition and learning tasks [2]. Many neuromorphic computers have architectures designed specifically to implement deep learning algorithms. In addition to these architecture changes, neuromorphic computers often use spikes to carry information. These "spiking neural networks" (SNNs) have very high computational power [3], [4], [5] and their hardware implementations have very low power consumption.

Spiking neuromorphic hardware may also require new physics and materials beyond traditional CMOS. We have previously proposed [6], [7] a hardware platform based on superconducting optoelectronic neurons (SOENs) that can achieve physical fanouts of one to one thousand [8], [9]. The platform utilizes the properties of superconductors to achieve energy efficiency similar to the human brain in terms of computations per second per watt, with operation energies as low as tens of attojoules per synapse event [7]. However, the proposed neurons are challenging to fabricate, and most approaches for training even on simple benchmark problems such as MNIST handwritten digit classification [10] require hundreds of neurons and tens of thousands of synapses. Meanwhile, it is usually not easy to take advantage of the full computational power of a spiking neuromorphic hardware platform when using these training approaches. In particular, training such networks effectively to perform useful tasks presents a major challenge. Here we propose using evolutionary optimization (EO) to take advantage of all of the parameters that we can control in this hardware platform to design smaller networks that are realistic to fabricate and test in the near term. We use the programming framework implemented by TENNLab [11] to simulate these circuits in a neuromorphic computation environment and provide guidance towards early SOEN design.

In this paper we propose a simplified circuit model for a SOEN neuron which we first simulate in LTSpice to obtain their electrical characteristics. We then implement a high level model of the simplified SOEN neuron/synapse circuits in the TENNLab evolutionary optimization programming framework [11] for neural networks. We use existing logic, control and classification applications already implemented in this framework to explore SOEN responses and performance while varying their design parameters. This technique provides feedback for the important parameters of these circuits and guides us towards early implementations of SOEN networks. Our goals in this study are to use EO to evaluate (1) if we can solve logic, control and classification tasks using this hardware platform (2) which of these circuit variations performs better and at which tasks, and (3) to investigate how the input and output encoding affects the circuit performance.

II. Hardware Platform

The spikes in this hardware platform consist of 20 ns photonic pulses, generated by a semiconductor LED [12]. The spikes are distributed by a network of photonic waveguides (shown in schematic in Fig. 1(a)) fabricated in deposited dielectrics [8], [13]. The photonic waveguides allow this hardware platform to reach very high physical fanout. A schematic of an individual neuron is shown in Fig. 1 (b). Pulses are received by downstream neurons at synapses (yellow). Each synapse converts the pulses from the optical to the electrical domain using a superconducting nanowire [13], with the

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

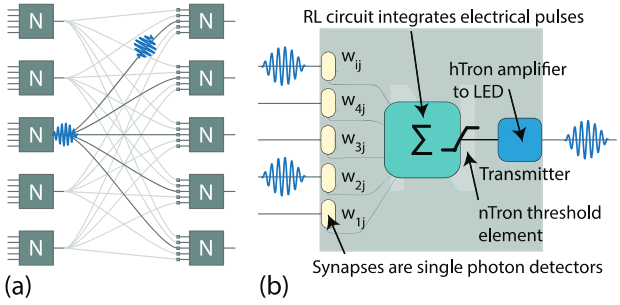


Fig. 1. (a) A network of neurons connected by photonic waveguides. (b) Schematic of an individual neuron.

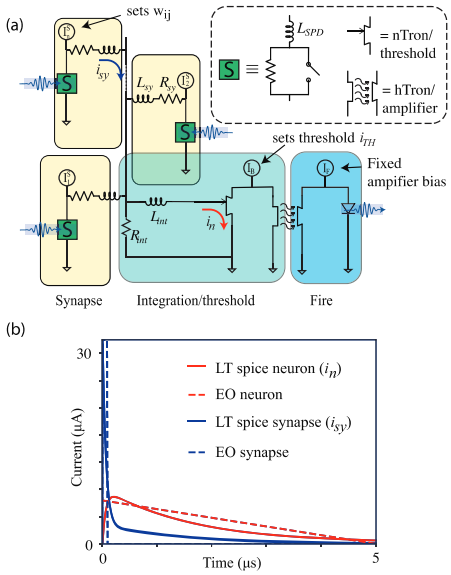


Fig. 2. (a) SOENs circuit with neuronal integration loop (SOEN1). The neuron fires a photonic pulse when $i_n > i_{TH}$. The equivalent circuit for the SNSPD is shown in the dashed box on the top right, as well as the nTron and hTron circuit symbols. (b) The current in the neuron integration loop/nTron (i_n) with two synapses firing (red) and the current ejected (i_{sy}) from the SNSPD (blue) when a synapse fires, simulated in LTspice (solid line) and in the EO programming framework (dashed line).

amplitude and duration of the electrical pulse determined by circuit properties described in the next section.

Varying levels of complexity are possible for SOEN design. In its simplest form, the amplitude of the electrical pulse from the detector synapse can be weighted (w_{ij}) and integrated with electrical pulses from other synapses on a thresholding unit, using a superconducting circuit element, the nTron [14] (Fig 1 (b)). Once the threshold value has been exceeded, the current in the integration loop triggers the transmitter circuit and discharges a photonic pulse. The transmitter circuit consists of an amplifier (hTron) and an LED coupled to an output waveguide (axon). This is the SOEN circuit design we will discuss in the remainder of this paper.

In more advanced versions of SOENs, the electrical pulse from the detectors is weighted by a high-efficiency superconducting Josephson junction (JJ) circuit [15], [16], which produces a number of single flux quantum (SFQ) pulses [17], [18], [19]. The number of SFQ pulses generated depends on the synaptic weight, which is in turn controlled by an input current. Further biological realism and processing capability are possible [15]. With as little as one photon from the upstream and downstream neuron, superconducting circuitry can adjust the current controlling the synaptic weight, thus implementing the biologically inspired learning mechanism spike-timing dependent plasticity (STDP). Similar circuitry can implement dendritic processing and metaplasticity. SFQ pulses from all of the synapses on a particular neuron can be integrated in a series of integration loops and compared to an adjustable threshold value.

At present, we have experimentally demonstrated the semiconductor LEDs, multiplanar waveguide routing, waveguide integrated superconducting detectors, and the superconducting electronics necessary for the thresholding and amplification elements. We have yet to demonstrate the synaptic Josephson junction circuits, although such technology is routinely implemented for voltage standards [20] and other superconducting circuits [21], [22], [23], [24]. For the earliest demonstrations, we seek to build a neuromorphic computer with simpler elements and direct electronic weight control in place of the more complicated JJ circuitry, albeit at lower energy efficiencies and with loss of some functionality. The advantage of these circuits is they provide us with a simple testbed with which we can demonstrate and develop this technology. Therefore the circuit described in this paper is a simplified version of the full circuit described in ref. [25]. The simplified circuit will enable us to prototype neurons and neural networks more rapidly.

III. Circuits

We start by considering two simplified SOEN variations, SOEN1 and SOEN2. SOEN1 is described in Fig. 2 (a). The neuron circuit can be divided into a number of synaptic circuits, a single neuronal integration loop and a transmitter circuit. Photonic spikes from upstream SOENs are received at the synapses of downstream neurons (indicated by an "S" in Fig. 2 (a)). Each synapse consists of a superconducting nanowire single photon detector (SNSPD) capable of detecting single photons. Single photon detection efficiency of >90% has been demonstrated with these detectors in the near-IR [26]. Synapses can be connected in parallel on a single neuron. When a photon is detected, the SNSPD ejects its current (i_{sy}), converting the photonic pulse to an electrical pulse. The equivalent circuit model for the SNSPD, used in the LTspice simulation, is shown in the dashed box in Fig. 2 (a) and consists of an inductor (L_{SPD}) in series with a switch in parallel with a 1 kΩ resistor. We model the detection of a photon on the SNSPD as opening the switch for 5 ns (with a rise time of 1 ns). The current pulse amplitude is set by the SNSPD bias current, and the actual pulse duration is given by the L/R time constant of the synapse, which depends on the

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

designed series resistance of the synapse, R_{sy}, the designed series inductance of the synapase, L_{sy}, the inductance of the detector, L_{SPD}, and the integration loop resistance, R_{int}. We refer to this detection of a photon and ejection of current as a synaptic firing event. Once a synapse has fired, it cannot fire again until the nanowire has returned to the superconducting state. In the present model, this time is 100 ns.

The current from the firing synapse is diverted to an accumulation element, which we refer to as the neuronal integration loop. When the synapse fires, it is stored as current (i_n) in the neuronal integration loop for a time (the neuron leak rate) given by the time constant of this loop (L_{int}/R_{int}). The amount of current added to this loop, and thus the synaptic weight, is varied by changing the bias current on the SNSPD in the synapse (labeled as I_i^S in Figs. 2 and 3). The synaptic pulse shown in Fig. 2 (b) is for a bias current of 15 μA, as would be present in an SNSPD made of MoSi or NbN. The current in the neuronal integration loop when two such synapses are firing at once is shown in red in the same figure. The neuronal integration loop has a leak rate of 200 kHz, leading to an integration time on the order of 5 μs, or 50 times longer than the SNSPD refractory period.

The thresholding element in the neuronal integration loop is an nTron superconducting current amplifier [14], which in this case is operated as a switch. The nTron will divert current to the transmitter circuit when the gate threshold, i_{th}, is exceeded, while i_{th} depends on the nTron bias current (I_{Bnt}), and can be varied dynamically by changing I_{Bnt}. The nTron will reset once it has been triggered, as the L/R time constant in the loop suddenly decreases due to the increase in R as the nTron gate turns from superconducting to normal, causing the current to leak out over a thousand times faster. The refractory period of the neuron is set by the reset time of the nTron, which is controlled by the new L/R time constant.

The transmitter circuit is triggered by a thresholding event to generate the photonic pulse referred to as a neuronal firing event. The transmitter circuit consists of a hTron amplifier, which transduces the small voltage generated by the nTron to the larger voltage necessary to turn on the LED. It operates by heating up a superconducting element in parallel with the LED, thus suddenly increasing the parallel resistance and diverting the current through the LED, as shown in Fig. 2 (a) between the threshold and fire boxes.

While the goal of this study is to determine if we can use evolutionary optimization to design neuromorphic systems using SOEN technology, we also consider whether evolutionary optimization can inform us on the minimum requirements for the basic SOEN design. In the SOEN circuits considered here, the neuronal integration loop serves as a short-term memory for spikes, but also reduces the amount of current passing through the nTron per synapse event. For reasonable values of the nTron threshold and synapse bias currents, a single synapse cannot trigger the neuron. Therefore we also consider the circuit shown in Fig. 3. In this circuit, the neuronal integration loop is omitted; this SOEN design has a much faster leak rate, defined by the L/R time constant of the SNSPDs. This leads

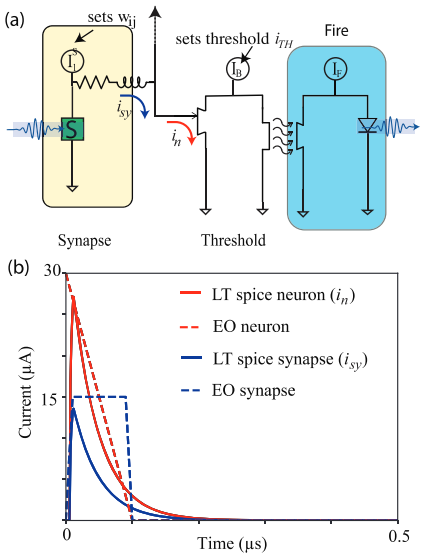


Fig. 3. (a) SOENs circuit without neuronal integration loop (SOEN2). The neuron will fire when $i_n > i_{TH}$. (b) Simulation of the current in nTron (i_n) with two synapses firing (red), the current ejected (i_{sy}) from the SNSPD (blue) when a synapse fires in LTspice and modeled in the EO programming framework.

to an integration time on the same order as the refractory period of the neuron and the synapse. The synapse pulse and the current through the nTron (thresholding element) when two synapses are firing at once for the no-integration circuit is shown in Fig. 3 (b). We reiterate that these are simple circuits that can be implemented in the superconducting optoelectronic technology in the near-term. Different combinations of the same circuit elements are also possible. We refer to all of these as SOEN circuits, with different parameter sets. More complex circuits may be developed as needed if a particular functionality is found to be critical, as we discuss in Section V. It is therefore one of the goals of this paper to determine the most useful circuit parameter sets for particular applications.

IV. Evolutionary Optimization

It is important to verify that we can use SOEN circuits in networks to perform various applications. Early in the development of a new technology it is difficult to design and fabricate systems of large numbers of neurons and synapses, as the infrastructure for emergent technologies is less mature. For example, even the relatively simple benchmark classification of MNIST handwritten digits is typically implemented in algorithms that use one input neuron per pixel (784 inputs) and ten output neurons (one per digit classification), in addition to order ten to one hundred hidden neurons. Moreover, implementations that simply map, for example, a convolutional neural network algorithm to a spiking hardware platform, do not take advantage of the dynamic aspect of SNNs, and the variety of ways that SNNs can encode information. It has been shown that an SNN can classify MNIST using a smaller

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

TABLE I
SIMPLIFIED MODEL FOR EO

EO parameters	SOEN1 (Fig. 2)	SOEN2 (Fig. 3)	SOEN3	SOEN4	SOEN5	SOEN6
neuron threshold	5-15 μA	5-15 μA	5-15 μA	5-15 μA	5-15 μA	5-15 μA
synapse weight	-3-3 μA	-15-15 μA	-3-3 μA	-15-15 μA	-15-15 μA	-15-15 μA
post-neuron delay	0	0	0	0	0-10 μs	0-10μs
Set parameters						
synapse delay	50 ns	"	"	"	"	"
neuron leak time	5 μs	0.1 μs	0.1 μs	5 μs	0.1 μs	5 μs
neuron refractory period	100 ns	"	"	"	"	"
synapse refractory period	100 ns	"	"	"	"	"

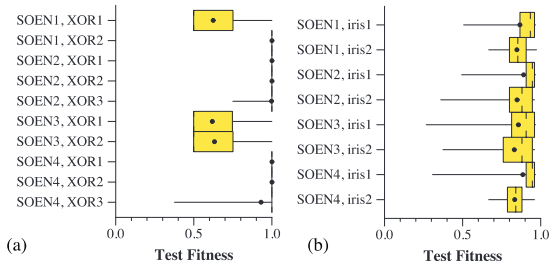


Fig. 4. (a) Statistics of testing fitness of evolved networks at XOR benchmark tests. XOR1 = (ppb = 1, W = 1), XOR2 = (ppb = 10, W = 1), XOR3 = (ppb = 10, W = 2), where W = number of bits, ppb = pulses per bin (see text). (b) Statistics for iris classification task (Iris1, ppb = 1, Iris2, ppb = 10).

TABLE II
XOR RESULTS (MAX, MEDIAN)

Model	XOR1	XOR2	XOR3
SOEN1	1.0, 0.5	1.0, 1.0	-
SOEN2	1.0, 1.0	1.0, 1.0	1.0, 1.0
SOEN3	1.0, 0.5	1.0, 0.5	-
SOEN4	1.0, 1.0	1.0, 1.0	1.0, 1.0

number of inputs [27], [28] by partially encoding the image in time. In Ref. [27], design of these relatively small networks was done by evolutionary optimization.

Motivated by the preceding arguments, in this paper we choose EO to design simple neural networks in the superconducting opto-electronic hardware platform. This allows us to take advantage of the complexity of dynamic spiking networks and to generate relatively small networks to solve particular tasks. EO has the additional advantages that it can use any of the parameters of this hardware, and will train the architecture (i.e. size and connectivity) of the network in addition to the parameters.

Table 1 shows the parameter sets we have tested for the SOENs model implemented in the TENNLab EO programming framework [11]. EO parameters are subject to mutation during training, while set parameters are defined in the model. Values are based on the LTspice simulations and realistic physical parameters, and are implemented as a first step towards a comprehensive model to design networks. SOEN1 and SOEN2 are also shown as the dashed lines in Figs. 2 (b) and 3 (b), and correspond most directly to the circuits described in the previous sections. However, there is flexibility in the design of these circuits, and therefore we also investigated other parameter combinations. For example, SOEN1 has a maximum synapse weight that is lower than the minimum neuron threshold, and therefore a single synapse cannot trigger a single neuron. Meanwhile, SOEN2 has a synapse weight range that will allow a single synapse to trigger a neuron, but with 50 times shorter integration time. Therefore, we have also implemented SOEN3 and SOEN4, to investigate whether differences between SOEN1 and SOEN2 are due to the synapse weight range or the integration time. We have not yet designed circuits with the SOEN3 and SOEN4 parameter sets, but rather we are using these parameter sets to investigate the importance of specific circuit parameters. In addition, if more parameters (e.g. synapse delay) are needed to solve tasks, more complex circuits may be designed for these tasks. Delay circuits and fixed electrical delays are possible to design using superconducting circuit elements. We therefore added post-neuron delays to the SOEN2 and SOEN4 parameter sets (generating SOEN5 and SOEN6), to investigate the effect of this additional parameter on the ability to evolve networks to perform various tasks.

To begin, we solved one of the TENNLab benchmark tasks, W-bit XOR. The task has been described in detail in Ref. [11]. We use two inputs per bit. A one bit XOR problem therefore requires four inputs. An "input" is a neuron to which current can be directly applied by the user or application, and serves as an initial transition between electrical inputs and photonic pulses. Depending on the magnitude of the electrical input a photonic pulse may or may not be produced by the input neuron, the magnitude of the photonic pulses are constant. Here we encode the input values in two different ways. In the first "no-rate-encoding" method a single pulse in one input neuron represents a one, a single pulse in the other input neuron represents a zero, i.e. pulse per bin is one (ppb = 1). In the second "rate-encoding" input method, 10 pulses are applied to the associated input neuron in the network (ppb = 10). In a digital problem such as XOR, "rate-encoding" is simply equivalent to adding redundancy to the inputs. For the rate-encoded inputs, there is an interval of 3 time steps between pulses, where a time step in the application translates to 1 μs for the device. In both cases, the applied input current is high enough that the input neuron will always fire a pulse. We run this task for one bit XOR, with no-rate-encoding (XOR1) and rate-encoding (XOR2) using the first four SOEN EO models described in Figs. 2, 3 and Table 1. Once the bits are pulsed into the neuromorphic device, the device runs for a set interval of time, during which output pulses are counted. The output that pulses more decides the answer. We also ran 2 bit XOR with rate-encoding (XOR3) for SOEN2 and SOEN4, which were overall the highest performing parameter sets.

Networks are initialized with the correct number of input neurons, output neurons and with one synaptic connection per neuron. In each epoch of an EO run, we run fitness calculations on a population of 1000 networks solving 200 XOR problems. The networks in the next epoch are generated based on the fitness results [29], and an algorithm that varies the topology

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

TABLE III
IRIS RESULTS (MAX, MEDIAN)

Model	Iris1	Iris2
SOEN1	0.96, 0.93	0.97, 0.85
SOEN2	0.96, 0.90	0.96, 0.88
SOEN3	0.96, 0.92	0.96, 0.88
SOEN4	0.96, 0.95	0.93, 0.79

of the networks as well as the "EO parameters" in Table I. Each EO run proceeds in as many epochs as it can complete in 5 hours (or until a fitness of 1 is reached). For each SOEN (model) and XOR (application) parameter set we perform 100 EO runs, generating 100 networks. These networks are then evaluated on five sets of 10,000 testing problems. The plots in Fig. 4 (a) show the testing fitness value statistics for the generated networks. The edges of the yellow box are the first and third quartiles of the data, the horizontal line is the median, the dot is the average, and the whisker extends to the max and min values of the data. Table II also summarizes the results for the best and median network fitness for each of the problems for both models. We find that while all of the models can generate networks with a fitness of 1, they do not all do so with equal likelihood, as can been seen from the fact that SOEN1 and SOEN3 have lower median fitnesses. In both SOEN1 and SOEN3, a single synapse cannot add enough current to the neuronal integration loop to trigger a neuron, due to the available nTron threshold range. It also appears that the added redundancy of the rate-encoding helps, in particular for the SOEN1 parameter set, which has a long integration time and performs much better for XOR2 (rate encoding) than XOR1.

We next investigate the performance of the circuits at the iris classification task from the UCI Machine Learning Repository [30], using the methods described in detail in [31]. In this task irises are classified as one of three types based on four physical measurements, the database consists of the measurement data labeled with the class. We use the same circuit and input variants as in the XOR problem, with four network inputs for the four iris characteristics. Iris1 therefore has ppb $= 1$ and iris2 ppb $= 10$. Each EO run is allowed to run for five hours and we report fitness in Fig. 4 (b). In this case we find very little variation, although the rate encoding tends to yield lower median network fitnesses. Overall, the XOR and iris problems seem to be too easy to solve, yielding little difference in the fitness of the evolved networks, although there are differences in the time and number of epochs it takes to arrive at these networks.

We next investigate the performance of SOEN neurons for a benchmark control task, balancing a pole on a cart. In this task, the network must keep a pole balanced on a cart as described in detail in Ref. [31]. The network can apply a fixed force to push the cart to the left or to the right to keep the pole balanced. The network receives values for the position and velocity of the cart and the angular position and velocity of the pole (four variables total). Each of these variables is binned

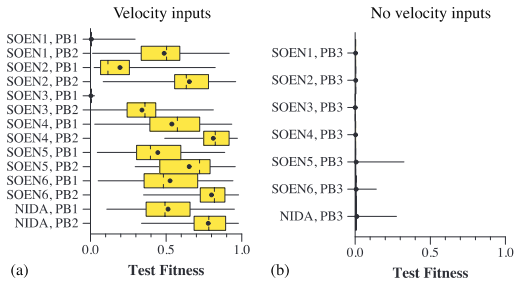


Fig. 5. (a) Statistics of testing fitness of evolved networks at a polebalance control task using position and velocity inputs. PB1 = (ppb = 1), PB2 = (ppb = 10) (b) Statistics of testing fitness at polebalance control task with only position inputs (ppb = 10). NIDA = TENNLab software model, SOEN parameter sets described in Table I.

into two categories, left and right. Thus, polebalance networks have a total of 8 input neurons. The current state of game is input to the network with activity on four of the input neurons, either with a single pulse (PB1) or ten pulses separated by 3 time steps (PB2). In PB1, the amplitude of the current pulse is allowed to vary, and this represents the value of the particular variable, for example cart position. The maximum amplitude will always cause the input neuron to fire, while intermediate amplitudes may or may not cause the input neuron to fire, depending on the evolved neuron threshold. In PB2, pulses are digital and the number of pulses represents the value. For example, if the cart position x can vary between -3 and 3, a value of $x = 1.2$ will cause 4 pulses to fire in the second bin, while a value of -1.2 will cause 4 pulses to fire in the first bin. In this case, an input current pulse will always cause an input neuron to fire. The network has two outputs, which effectively vote on which direction to apply the force.

We ran the SOEN model for all six parameter sets and the TENNLab NIDA model [11] at the PB1 and PB2 tasks. The results of the tests are shown in Fig. 5 (a) and Table IV. All of the models performed worse for PB1. However, SOEN1 and SOEN3 had the most difficulty solving this task, again showing the importance of a single synapse triggering a neuron. This may prove possible to overcome by the selection of suitable input parameters, EO mutation weights or initial conditions. SOEN4, which has a long integration time and a high maximum synapse weight, performed the best, with SOEN2, which has a shorter integration time and the same maximum synapse weight performing second best. This shows that the addition of an integration time is beneficial in this task. Comparing SOEN2/SOEN5 or SOEN4/SOEN6 pairs, which are the same except for the addition of a post-neuron delay, we find that there is a small improvement with the addition of this parameter.

In principle, we should be able to develop networks that do not require velocity as an input, but rather use only the position of the cart and pole to estimate the velocity within the network. Such networks would "remember" previous positions and infer the velocity information. We test the ability of the

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

TABLE IV
POLE BALANCER RESULTS (MAX, MEDIAN)

Model	PB1	PB2	PB3
SOEN1	0.30, 0.002	0.92, 0.5	0.004, 0.009
SOEN2	0.82, 0.11	0.96, 0.63	0.004, 0.004
SOEN3	0.03, 0.002	0.81, 0.36	0.002, 0.001
SOEN4	0.93, 0.57	0.97, 0.82	0.004, 0.003
SOEN5	0.89, 40	0.96, 0.72	0.32, 0.003
SOEN6	0.94, 0.48	0.98, 0.82	0.14, 0.003
NIDA	0.95, 0.49	0.98, 0.78	0.27, 0.004

SOENs models to perform this task with only the positions as inputs. In this case, there are only two input bins for cart position and two for pole angular position (for a total of four input neurons), and we always use rate encoding with up to ten pulses per bin. We first use the TENNLab NIDA model [11] and show that it is capable of solving the problem. We find that NIDA has more difficulty than in the case with velocity as an input, evolving many low fitness networks (the average, median and first and third quartiles are all close to zero), but is still able to produce a few higher fitness networks. We next test it for the SOEN parameter sets in the previous test. Even SOEN2 and SOEN4 were not able to balance the pole without the velocity inputs (using the same time limit for the each EO run as in the previous tests). The main difference between the SOEN model with parameter sets 1-4 and NIDA as implemented here, is that NIDA has no leak (so it holds charge forever) and second, the synapses have a variable and evolvable delay. This indicates that we may need to implement such a variable synapse delay, or some other surrogate for it, in these circuits in order to perform certain tasks. SOEN5 and SOEN6 have the same parameter sets as SOEN2 and SOEN4, but with the addition of an evolvable post-neuron delay. We find that both models perform similarly to NIDA, evolving a few networks with reasonable fitnesses. This suggests that while variable delays may not be important for some tasks (e.g. the polebalance with velocity inputs) it may be worthwhile for other tasks (e.g. the polebalance without velocity inputs). Therefore it is likely worthwhile to design some variable delay circuits, but it may be only necessary to implement them after each neuron, rather than an element per synapse.

V. DISCUSSION AND FUTURE WORK

We were able to design networks for logic, control and classification using the SOEN model in the TENNLab programing framework. We investigated various parameter sets for the SOEN model to help inform the designs of our early circuits. We found that a larger synapse weight range, such that it is possible for a single synapse to trigger a single neuron, works better for every application we tested, and was more important than a long integration time. However, secondary to the synaptic weight range, a longer neuron integration time does seem to add some benefit. Because of this, we may work to engineer an nTron with lower threshold values or SNSPDs with higher threshold currents, e.g. using thicker superconducting films. Alternatively, we could investigate changing the EO starting parameters such that neurons start with a large number of synaptic connections We also find evidence that the presence of time-dependent variables, such as synapse delay may be critical for solving certain tasks. In this case, we could only balance a pole without velocity information (requiring memory of past events) if the model included an evolvable timing delay.

One important lesson for SOEN circuit design using this programming framework is that it is important to use the right input encoding for a particular device. For example, in the case of pole balancer, the SOEN circuit would not solve the task without rate encoding for some parameter sets. The encoding that is used will effect the networks generated, and may effect the outcome of the optimization (i.e. whether or not it is possible to find networks to solve a particular task).

There are many further questions that arise from this study. Intuitively, we expect that integration is more important in applications where inputs arrive to the network stochastically. In future work we will investigate whether we find this to be the case. We also plan to investigate whether it is possible to use other time-dependent variables in place of the synapse delay in SOEN circuits. It would also be interesting to investigate if a variable refractory period or diverse synaptic leak rates could be used to similar effect as neuron and synapse delay, and for which tasks this is important. If these features are crucial for many tasks, then we will need to include them in future SOEN circuit models.

This is an early model that we hope to use as a stepping stone towards more concrete designs and, ultimately, experimental demonstrations. Future work will be to verify these models against network simulations using low level software such as Verilog A. We will also verify the designed networks against experimentally fabricated networks. In this case, we will also see if further optimization can be performed on the hardware itself. Finally, we are also interested in extending the framework to implement an energy and area evaluation of the produced networks. Ultimately this could allow the reward of networks that use lower energy and minimize total wiring.

REFERENCES

[1] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, "A Survey of Neuromorphic Computing and Neural Networks in Hardware," may 2017. [Online]. Available: http://arxiv.org/abs/1705.06963

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, may 2015. [Online]. Available: http://www.nature.com/articles/nature14539

[3] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, dec 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608097000117

[4] W. Gerstner and W. Kistler, *Spiking neuron models*, 1st ed. Cambridge: Cambridge University Press, 2002.

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

[5] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, jan 2018. [Online]. Available: http://ieeexplore.ieee.org/document/8259423/

[6] J. Shainline, S. Buckley, R. Mirin, and S. Nam, "Superconducting optoelectronic circuits for neuromorphic computing," *Phys. Rev. App.*, vol. 7, p. 034013, 2017.

[7] J. M. Shainline, S. M. Buckley, A. N. McCaughan, J. Chiles, R. P. Mirin, and S. W. Nam, "Superconducting Optoelectronic Neurons I: General Principles," may 2018. [Online]. Available: http://arxiv.org/abs/1805.01929

[8] J. Chiles, S. Buckley, N. Nader, S. W. Nam, R. P. Mirin, and J. M. Shainline, "Multi-planar amorphous silicon photonics with compact interplanar couplers, cross talk mitigation, and low crossing loss," *APL Photonics*, vol. 2, no. 11, p. 116101, nov 2017. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.5000384

[9] J. Chiles, S. M. Buckley, S. W. Nam, R. P. Mirin, and J. M. Shainline, "Design, fabrication, and metrology of 10 100 multi-planar integrated photonic routing manifolds for neural networks," *APL Photonics*, vol. 3, no. 10, p. 106101, oct 2018. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.5039641

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Online]. Available: http://ieeexplore.ieee.org/document/726791/

[11] J. S. Plank, G. S. Rose, M. E. Dean, C. D. Schuman, and N. C. Cady, "A Unified Hardware/Software Co-Design Framework for Neuromorphic Computing Devices and Applications," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*. IEEE, nov 2017, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/document/8123655/

[12] S. Buckley, J. Chiles, A. N. McCaughan, G. Moody, K. L. Silverman, M. J. Stevens, R. P. Mirin, S. W. Nam, and J. M. Shainline, "All-silicon light-emitting diodes waveguide-integrated with superconducting single-photon detectors," *Applied Physics Letters*, vol. 111, no. 14, p. 141101, oct 2017. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.4994692

[13] J. Shainline, S. Buckley, N. Nader, C. Gentry, K. Cossel, J. Cleary, M. Popović, N. Newbury, S. Nam, and R. Mirin, "Room-temperature-deposited dielectrics and superconductors for integrated photonics," *Opt. Express*, p. 10322, 2017.

[14] A. N. McCaughan and K. K. Berggren, "A Superconducting-Nanowire Three-Terminal Electrothermal Device," *Nano Letters*, vol. 14, no. 10, pp. 5748–5753, oct 2014. [Online]. Available: http://pubs.acs.org/doi/10.1021/nl502629x

[15] J. M. Shainline, A. N. McCaughan, S. M. Buckley, C. A. Donnelly, M. Castellanos-Beltran, M. L. Schneider, R. P. Mirin, and S. W. Nam, "Superconducting Optoelectronic Neurons III: Synaptic Plasticity," may 2018. [Online]. Available: http://arxiv.org/abs/1805.01937

[16] J. M. Shainline, S. M. Buckley, A. N. McCaughan, M. Castellanos-Beltran, C. A. Donnelly, M. L. Schneider, R. P. Mirin, and S. W. Nam, "Superconducting Optoelectronic Neurons II: Receiver Circuits," may 2018. [Online]. Available: http://arxiv.org/abs/1805.02599

[17] M. Tinkham, *Introduction to Superconductivity*, 2nd ed. Dover, 1996.

[18] T. V. Duzer and C. Turner, *Principles of superconductive devices and circuits*, 2nd ed. USA: Prentice Hall, 1998.

[19] A. M. Kadin, *Introduction to superconducting circuits*, 1st ed. USA: John Wiley and Sons, 1999.

[20] S. P. Benz and C. A. Hamilton, "A pulsedriven programmable Josephson voltage standard," *Applied Physics Letters*, vol. 68, no. 22, p. 3171, jun 1998. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.115814

[21] K. Likharev, "Superconductor digital electronics," *Physica C*, vol. 482, p. 6, 2012.

[22] N. Takeuchi, D. Ozawa, Y. Yamanashi, and N. Yoskikawa, "An adiabatic quantum flux parametron as an ultra-low-power logic device," *Superconductor Science and Technology*, vol. 1126, p. 1, 2013.

[23] Q. Herr, A. Herr, O. Oberg, and A. Ioannidis, "Ultra-low-power superconductor logic," *J. Appl. Phys.*, vol. 109, p. 103903, 2011.

[24] G. Wendin, "Quantum information processing with superconducting circuits: a review," *Rep. Prog. Phys.*, vol. 80, p. 106001, 2017.

[25] J. M. Shainline, A. N. McCaughan, S. M. Buckley, R. P. Mirin, and S. W. Nam, "Superconducting Optoelectronic Neurons IV: Transmitter Circuits," may 2018. [Online]. Available: http://arxiv.org/abs/1805.01941

[26] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Detecting single infrared photons with 93% system efficiency," *Nature Photonics*, vol. 7, no. 3, pp. 210–214, feb 2013. [Online]. Available: http://dx.doi.org/10.1038/nphoton.2013.13

[27] C. D. Schuman, J. D. Birdwell, and M. Dean, "Neuroscience-inspired inspired dynamic architectures," in *Proceedings of the 2014 Biomedical Sciences and Engineering Conference*. IEEE, may 2014, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/6867735/

[28] C.-K. Lin, A. Wild, G. N. Chinya, Y. Cao, M. Davies, D. M. Lavery, and H. Wang, "Programming Spiking Neural Networks on Intel's Loihi," *Computer*, vol. 51, no. 3, pp. 52–61, mar 2018. [Online]. Available: http://ieeexplore.ieee.org/document/8303802/

[29] C. D. Schuman, J. S. Plank, A. Disney, and J. Reynolds, "An evolutionary optimization framework for neural networks and neuromorphic architectures," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, jul 2016, pp. 145–154. [Online]. Available: http://ieeexplore.ieee.org/document/7727192/

[30] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[31] C. D. Schuman, A. Disney, S. P. Singh, G. Bruer, J. P. Mitchell, A. Klibisz, and J. S. Plank, "Parallel Evolutionary Optimization for Neuromorphic Network Training," in *2016 2nd Workshop on Machine Learning in HPC Environments (MLHPC)*. IEEE, nov 2016, pp. 36–46. [Online]. Available: http://ieeexplore.ieee.org/document/7835813/

Buckley, Sonia; McCaughan, Adam; Chiles, Jeffrey; Mirin, Richard; Nam, Sae Woo; Shainline, Jeffrey.
"Design of superconducting optoelectronic networks for neuromorphic computing."
Paper presented at 2018 IEEE International Conference on Rebooting Computing, Tysons, VA, United States. November 7, 2018 - November 9, 2018.

An Ultra-fast Multi-level $MoTe_2$-based RRAM

F. Zhang[1], H. Zhang[2,3], P.R. Shrestha[2,3], Y. Zhu[1], K. Maize[1], S. Krylyuk[2,3], A. Shakouri[1], J.P. Campbell[3], K.P. Cheung[3], L.A. Bendersky[3], A.V. Davydov[3] and J. Appenzeller[1*]

[1]Purdue University, West Lafayette, IN, USA, *email: appenzeller@purdue.edu,

[2]Theiss Research, Inc., La Jolla, CA, USA, [3]National Institute of Standards and Technology, Gaithersburg, MD, USA,

***Abstract* —** We report multi-level $MoTe_2$-based resistive random-access memory (RRAM) devices with switching speeds of **less than 5 ns** due to an electric-field induced 2H to $2H_d$ phase transition. Different from conventional RRAM devices based on ionic migration, the $MoTe_2$-based RRAMs offer intrinsically better reliability and control. In comparison to phase change memory (PCM)-based devices that operate based on a change between an amorphous and a crystalline structure, our $MoTe_2$-based RRAM devices allow faster switching due to a transition between *two* crystalline states. Moreover, utilization of atomically thin 2D materials allows for aggressive scaling and high-performance flexible electronics applications. Multi-level stable states and synaptic devices were realized in this work, and operation of the devices in their low-resistive, high-resistive and intrinsic states was quantitatively described by a novel model.

I. Introduction

RRAM has promises of being an emerging technology due to its potential scalability, high operation speed, high endurance and ease of process flow. However, reliable and repeatable operation is a potential challenge in future applications since switching involves the uncontrollable motion of individual atoms. In this work, we present a new switching mechanism for $MoTe_2$-based RRAM. An electric field induces the phase transition from the stable semiconducting 2H phase to a more conductive $2H_d$ phase, which provides a potential path towards better stability. The newly formed $2H_d$ is structurally close to the 2H phase, which holds the promise for faster switching if compared with the significant migration of ions in conventional RRAM [1] or the amorphous-to-crystalline transition in PCM [2] (see Fig.1). Initial pulse measurements show impressive switching speed of less than 5 ns. Moreover, multi-level states can be programmed into the devices by applying proper set/reset voltages, which allows to gradually changing the device resistance with multiple pulses, creating a "synaptic device".

II. Switching in $MoTe_2$ RRAM Devices

Fig. 1 illustrates the advantages of this new type $MoTe_2$-based RRAM as compared to conventional RRAM and PCM due to switching by an electric-field induced phase transition. Fig. 2(a) shows schematically a vertical $MoTe_2$ device. First, a bottom electrode Ti/Au (10 nm/25 nm) was deposited onto a 90 nm silicon dioxide (SiO_2) layer covering a highly doped silicon wafer. Next, $MoTe_2$ (2D Semiconductor) layers were exfoliated onto this electrode using standard scotch tape techniques, followed by thermal evaporation of 55 nm SiO_2 insulating layer. The device fabrication was finished by the deposition of a Ti/Ni (35 nm/50 nm) top electrode. Different from previously reported CVD grown 2D material based RRAM devices [3,4] whose operation is mediated by uncontrollable defects/grain boundaries in the device structure, our active material is a single-crystalline layer, where the observed RRAM behavior is due to the intrinsic properties of $MoTe_2$. Fig. 2(b) shows an AFM image of a $MoTe_2$ vertical device. The active region is about 0.1 μm^2. Fig. 3(a) displays I-V curves of a pristine device, the device forming process and successive cycling through its high resistive state (HRS) and low resistive state (LRS). Stable and reproducible bipolar RRAM behavior was observed. Fig. 3(b) shows the I-V curves for $MoTe_2$ devices with different layer thicknesses, and Fig. 3(c) summarizes how the forming and set voltages scale with the $MoTe_2$ layer thickness.

Thermoreflection microscope images were acquired to map the location of the filament on the device after forming. Surface temperature maps with 50 mK temperature resolution and submicron (diffraction limited) spatial resolution can be acquired within a few minutes [5]. Fig. 4(a) and (b) show self-heating hotspots on the nickel electrode surface superimposed with optical images for two representative devices, indicating the position of the filaments. The occurrence of a single hotspot for each device with characteristic hotspot full-width-at-half-maximum of ~ 200 nm is consistent with joule self-heating from a source as small as a $MoTe_2$ filament. In six out of eight devices imaged the hotspot was located at the edge or corner of the active region as can be seen in Fig. 4(a) and (b). We speculate that this preferential occurrence is a result of stronger electric fields at "sharp" topological features due to patterning during the fabrication that enhances the filament formation. Fig. 4(c) shows the calibrated temperature change map for the hotspot in device (b). The detected temperature change on the filament portion is ~ 15 K.

In order to understand the filament formation mechanism in $MoTe_2$, scanning transmission electron microscopy (STEM) of cross-sectional samples was utilized. As Fig. 5 shows, in the LRS, a distorted $2H_d$ phase was identified in the regions extending vertically throughout the $MoTe_2$ layer. The $2H_d$ phase was identified as a distorted modification of the 2H structure – a transient state with atoms displaced to the sites of a lower symmetry, but still within atomic arrangements of the 2H structure. A detailed analysis of the structure can be found in ref. [6]. Fig. 6 shows an energy dispersive spectrometry (EDS) scan along the filament region of a device in its LRS. Almost no Ti and Au signals were detected within the filament, which – in particular when also considering the unavoidable ion-milling contamination during FIB sample preparation – implies that the switching mechanism is *not* related to the migration of metal ions. To further confirm this point, graphene was used to replace the metal top and bottom electrodes. Fig. 7 shows the "typical" bipolar RRAM behavior

Davydov, Albert; Bendersky, Leonid; Krylyuk, Sergiy; Zhang, Huairuo; Zhang, Feng; Appenzeller, Joerg; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"An Ultra-fast Multi-level MoTe2-based RRAM."
Paper presented at 2018 IEEE International Electron Device Meeting, IEDM 2018, San Francisco, CA, United States. December 1, 2018 - December 5, 2018.

observed here in a Graphene-$MoTe_2$-Graphene device. Note that this is the first demonstration of an entirely 2D materials-based RRAM. Based on previous studies [7], graphene is a good diffusion barrier for metal ions. This 2D RRAM excludes completely the possibility of migration of metal ions as a source for the resistive switching observed by us. Based on these results, an electric-field induced phase transition from 2H to a more conductive $2H_d$ state is believed to be responsible for the RRAM behavior in vertical $MoTe_2$ devices.

III. A Physical Model

To fully understand the vertical transport through the pristine, LRS and HRS states in $MoTe_2$ devices and to explore the properties of the new $2H_d$ phase, a physical model was constructed. The barrier height Φ_{2H} and Φ_{2Hd} shown in figure 8(a) were extracted by utilizing the numerical model from ref. [8]. In this model, two different transport mechanisms are considered: thermal diffusion at low voltages and Fowler-Nordheim (FN) tunneling at higher voltages, as illustrated for a pristine data set in Fig. 8(b). In our model the 2H phase has a larger barrier height than the $2H_d$ phase as evident when their current values in the low voltage range are compared. An excellent fit can be obtained for the pristine I-V characteristics (Fig. 8(b)), as well as for the LRS and HRS (Fig. 8(c)) by employing the band diagrams and parameters shown in Fig. 8(a). In the pristine state, the $MoTe_2$ is in its 2H phase with a large barrier height of $\Phi_{2H} = 0.38$ eV. On the other hand, in the LRS, a filament of $2H_d$ was created through the setting process with an extracted barrier height of $\Phi_{2Hd} \approx 0.07$ eV. The HRS is characterized by formation of the $2H/2H_d$ heterojunction due to rupture of the $2H_d$ filament during the reset process. The thickness of newly formed 2H segment in the filament can be estimated to be ~ 1.8nm. Thus, the simulation results are consistent with the notion that a new semiconducting $2H_d$ state with a smaller barrier height is formed during the set process that is responsible for the higher conductivity of the LRS compared with the HRS.

IV. Performance Study and Pulse Measurements

A. Performance study

Fig. 9 illustrates the pulse switching behavior in $MoTe_2$ based RRAM devices. The pulse width is 80 μs. By applying set/reset pulses, the device can switch between the LRS and HRS. Fig. 9(c) shows the read out current per cycle in the LRS and HRS. Fig. 9(d) is a retention measurement. All performance studies indicate stable and reproducible RRAM behavior.

B. Pulse measurements

To test the switching speed of our $MoTe_2$ RRAM devices, the experimental measurement setup shown in the inset of Fig. 10(b) was utilized. The current through the device was measured using a 50 Ω termination at the oscilloscope [9]. Note that no current compliance was used in this setup. The switching was controlled by varying the pulse width or pulse amplitude. A current and voltage versus time (t) plot of one such SET operation is shown in Fig. 10(a). The full width at half max (FWHM) of the applied voltage pulse is 5 ns. Fig. 10(b) shows the same data as in Fig. 10(a) but plotted as the current versus voltage. This clear change in resistance during the 5 ns voltage pulse is further evidence that the switching speed in $MoTe_2$ based RRAM is less than 5 ns.

The devices were reset (switched OFF) using negative pulses. Multiple pulses were required for a gradual reset process. Fig. 10(c) shows 10 pulses used to reset the device. Current versus voltage plots of cycle 1 (1^{st} pulse), cycle 5 (5^{th} pulse) and cycle 10 (10^{th} pulse) are shown in Fig. 10(d). The gradual resistance change is a desirable feature for neuromorphic computing. Fig. 11 shows the characteristics of a device that was programmed by a series of positive pulses (1.1 V, 80 μs) followed by a series of negative voltage pulses (-1.2 V, 80 μs). The resistance of the $MoTe_2$ device gradually decreased and increased, which is similar to the potentiation and depression of biological synapses.

By carefully tuning the set/reset voltages, multi-level states can be programmed into the devices. Fig. 12(a) shows stable resistive states after various short (80 μs) and long (560 μs) voltage pulses that were read at a 0.2 V level. Long pulses result in a more substantial change (training) of the resistive state of the system. Fig. 12(b) shows the switch on/off behavior in each state. All the pulse measurements hint at an additional application space of this class of vertical TMD devices in the realm of neuromorphic computing.

Acknowledgment

This work was supported in part by the Semiconductor Research Corporation (SRC) as the NEWLIMITS Center and NIST through award number 70NANB17H041. H.Z. acknowledges support from the U.S. Department of Commerce, NIST under the financial assistance awards 70NANB15H025 and 70NANB17H249. S. K. acknowledges support from the U.S. Department of Commerce, NIST under the financial assistance award 70NANB18H155.

References

[1] H. S. P. Wong, *et al.*, "Metal-Oxide RRAM," *Proceedings of the IEEE,* vol. 100, pp. 1951-1970, 2012.

[2] H. S. P. Wong, *et al.*, "Phase Change Memory," *Proc. IEEE,* vol. 98, pp. 2201-2227, 2010.

[3] V. K. Sangwan, *et al.*, "Gate-tunable memristive phenomena mediated by grain boundaries in single-layer MoS_2," *Nat. Nanotechnol.,* vol. 10, pp. 403-406, 05//print 2015.

[4] F. M. Puglisi, *et al.*, "2D h-BN based RRAM devices," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 34.8.1-34.8.4.

[5] K. Maize, *et al.*, "Transient Thermal Imaging Using Thermoreflectance," in *2008 Twenty-fourth Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, 2008, pp. 55-58.

[6] F. Zhang, *et al.*, "Electric field induced semiconductor-to-metal phase transition in vertical MoTe2 and $Mo_{1-x}W_xTe_2$ devices," *arXiv preprint arXiv:1709.03835,* 2017.

[7] R. Mehta, *et al.*, "Transfer-free multi-layer graphene as a diffusion barrier," *Nanoscale,* vol. 9, pp. 1827-1833, 2017.

[8] Y. Zhu, *et al.*, "Vertical charge transport through transition metal dichalcogenides–a quantitative analysis," *Nanoscale,* vol. 9, pp. 19108-19113, 2017.

[9] P. R. Shrestha, *et al.*, "Compliance-free pulse forming of filamentary RRAM," *ECS Transactions*, vol. 75, pp. 81-92, 2016.

Davydov, Albert; Bendersky, Leonid; Krylyuk, Sergiy; Zhang, Huairuo; Zhang, Feng; Appenzeller, Joerg; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"An Ultra-fast Multi-level MoTe2-based RRAM."
Paper presented at 2018 IEEE International Electron Device Meeting, IEDM 2018, San Francisco, CA, United States. December 1, 2018 - December 5, 2018.

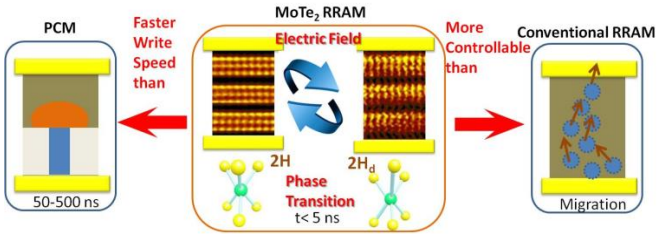


Fig. 1. Highlight features of the $MoTe_2$ based RRAM.

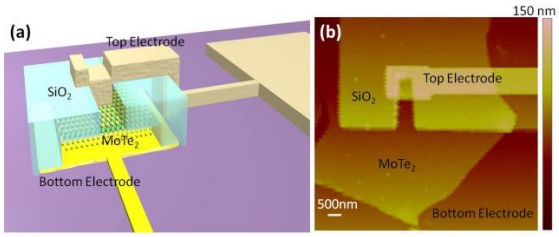


Fig. 2. (**a**) Schematic diagram of a vertical metal/$MoTe_2$/metal device. (**b**) AFM image of a vertical $MoTe_2$ device.

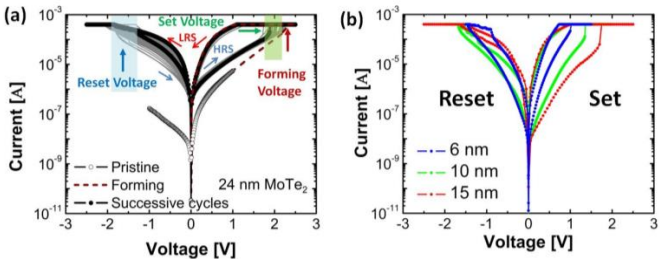


Fig. 3. (**a**) I-V curves of a 24 nm $MoTe_2$ device with an active area of 330 nm x 500 nm. 40 cycles are shown in the grey line curves. Current compliance is set to 400 μA. (**b**) I-V curves of vertical $MoTe_2$ RRAM devices from 6 nm, 10 nm, and 15 nm $MoTe_2$ layers with active areas of 502 nm x 360 nm, 522 nm x 330 nm and 500 nm x 330 nm respectively. (**c**) Forming/Set voltage values scale with the $MoTe_2$ thickness.

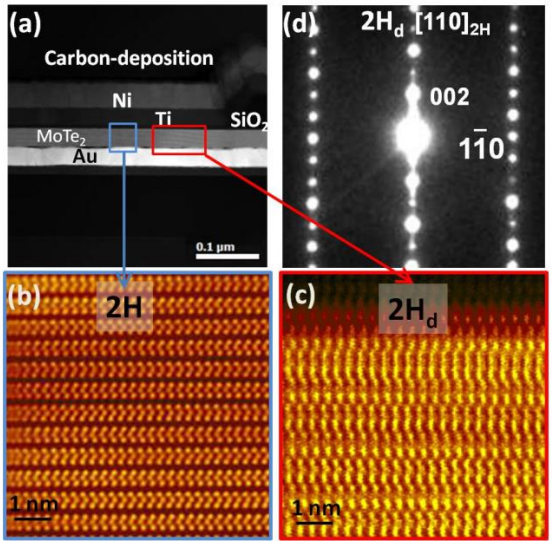


Fig.5. (**a**) HAADF-STEM image showing the cross-section of a $MoTe_2$ device. Higher magnification HAADF image from the region defined by the blue/red box in (**a**) showing (**b**) 2H and (**c**) a distorted structure ($2H_d$) taken along the $[110]_{2H}$ zone-axis. (**d**) Corresponding nano-beam diffraction pattern taken from the distorted $2H_d$ area.

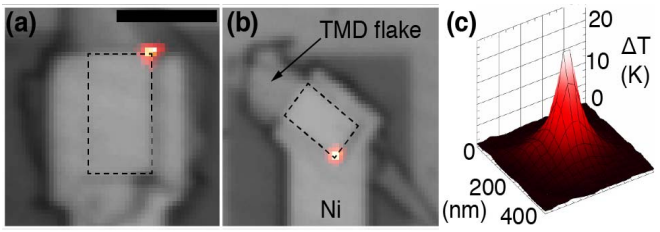


Fig. 4.(**a-b**) Thermoreflectance images showing the location of the filament. The scale bar is 5 μm. (**c**) The calibrated temperature change map for the filament in the device (**b**).

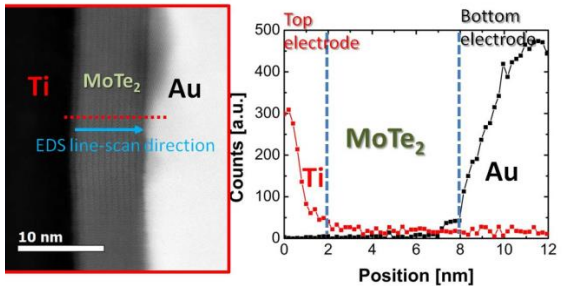


Fig. 6. EDS line-scan analysis of the filament region of a device in its LRS. Little Ti/Au signals were detected in the filament region due to ion-milling contamination during sample preparation.

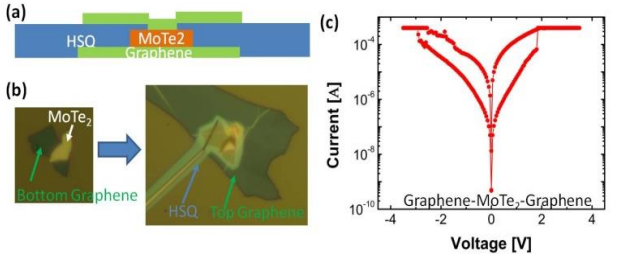


Fig. 7. (**a**) Schematic and (**b**) optical images of a Graphene-$MoTe_2$-Graphene device. (**c**) I-V curve of the RRAM device in (**b**).

Davydov, Albert; Bendersky, Leonid; Krylyuk, Sergiy; Zhang, Huairuo; Zhang, Feng; Appenzeller, Joerg; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"An Ultra-fast Multi-level MoTe2-based RRAM."
Paper presented at 2018 IEEE International Electron Device Meeting, IEDM 2018, San Francisco, CA, United States. December 1, 2018 - December 5, 2018.

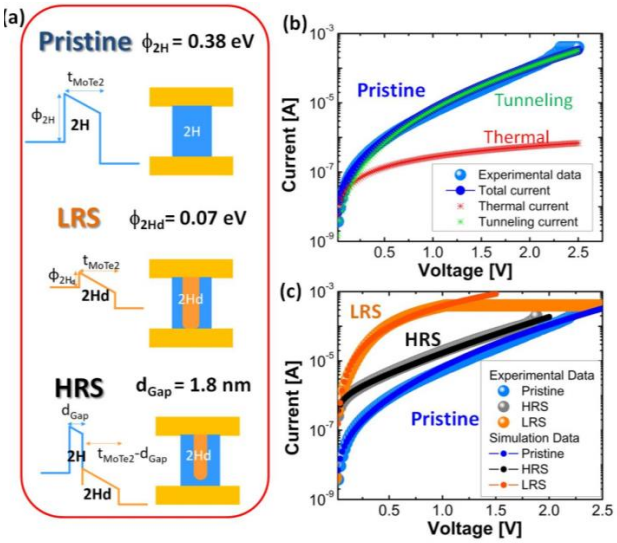


Fig. 8. (**a**) Schematic band diagrams of the RRAM in its various states. (**b**) Experimental and simulation data of a pristine device. (**c**) Experimental and simulation data for the LRS, HRS and pristine state.

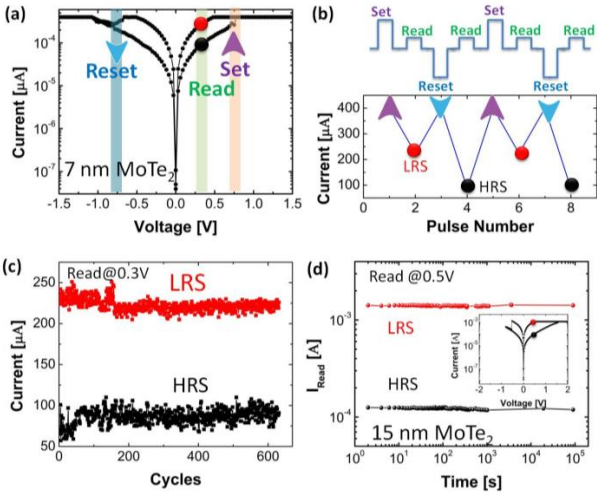


Fig. 9. (**a**) DC characteristic of a 7 nm thick $MoTe_2$ layer RRAM device. (**b**) shows the various current levels after pulse switching. (**c**) Current versus cycle in the LRS and HRS at a read voltage of 0.3 V. Current compliance is set to 400 μA. (**d**) Retention of the HRS and LRS for a 15 nm thick $MoTe_2$ RRAM device. Current compliance is set to 1.2 mA.

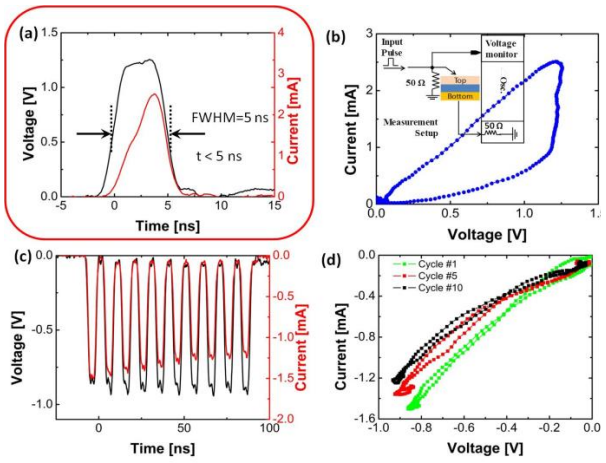


Fig. 10. (**a**) I/V vs. time plot showing switching of the device within a 5 ns voltage pulse. (**b**) I-V plot of the data shown in (**a**) to show the change in resistance during the applied pulse. The inset figure displays the experimental setup for the pulse measurement. (**c**) I/V vs. t plot of 10 reset pulses. (**d**) I-V plot of data plotted in (**c**) of pulses 1, 5 and 10.

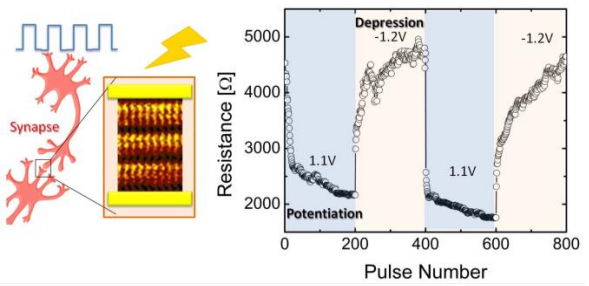


Fig. 11. Synaptic device from an $MoTe_2$ RRAM.

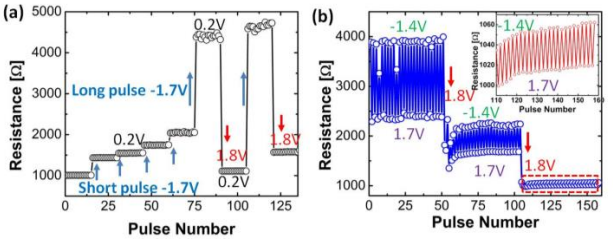


Fig. 12. (**a**) Multiple stable states of a device after various set and reset voltage pulses had been applied. Read out occurs at a voltage of 0.2 V. Every state is characterized by 15 subsequent read outs. A short pulse (80 μs) and longer pulse (560 μs) of a reset voltage of -1.7 V result in different changes of the resistance. (**b**) Multi-level characteristics of a vertical $MoTe_2$ device. Each level exhibits stable switch on/off at set/reset voltages of 1.7 V and -1.4 V respectively. A third state can be "dialed in" through yet another 1.8 V pulse. The inset figure is the zoom-in of the red dashed part.

Davydov, Albert; Bendersky, Leonid; Krylyuk, Sergiy; Zhang, Huairuo; Zhang, Feng; Appenzeller, Joerg; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"An Ultra-fast Multi-level MoTe2-based RRAM."
Paper presented at 2018 IEEE International Electron Device Meeting, IEDM 2018, San Francisco, CA, United States. December 1, 2018 - December 5, 2018.

Improving dielectric nano-resonator-based antireflection coatings for photovoltaics

Dongheon Ha*[a, b], Nikolai B. Zhitenev[a]
[a]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD, USA 20899; [b]Maryland Nanocenter, University of Maryland, College Park, MD, USA 20742

ABSTRACT

We demonstrate optical and electrical property enhancement of solar cells using a variety of dielectric nano-resonator array coatings. First, we study close-packed silicon dioxide (SiO_2) nano-resonator arrays on top of silicon (Si) and gallium arsenide (GaAs) solar cells. From macroscale measurements and calculations, we find that absorptivity of solar cells can be improved by 20 % due to the resonant couplings of excited whispering gallery modes and the thin-film antireflection effect. Next, we image photocurrent enhancement at the nanoscale via near-field scanning photocurrent microscopy (NSPM). Strong local photocurrent enhancement is observed over each nano-resonator at wavelengths corresponding to the whispering gallery mode excitation. Finally, for better optical coupling to solar cells, we explore hybrid nano-resonator arrays combining multiple materials such as silicon dioxide, silicon nitride, and titanium dioxide. Due to higher number of photonic modes within such hybrid coatings, absorptivity is enhanced by more than 30 % in a Si solar cell.

Keywords: Antireflection coatings, nanospheres, nanoresonators, whispering gallery modes, absorptivity enhancements, photocurrent enhancements, near-field scanning optical microscopy (NSOM), near-field scanning photocurrent microscopy (NSPM)

1. INTRODUCTION

Innovative nanoscale light management schemes have been introduced in order to make cost-effective photovoltaics (PVs) [1]-[10]. Among various approaches, light management using arrays of dielectric nano-resonators has been considered a promising route to significantly increase light absorption and power conversion efficiency of solar cells due to several positive aspects. This method does not require direct surface patterning of active materials, and therefore the open circuit voltage remains intact. Furthermore, this method is relatively cheap. However, there are some other factors that still need to be addressed to apply such light management technique to commercial PVs. First, as only macro-/microscale characterizations have been made so far, clear demonstration of nanoscale optical coupling and photocarrier collection is still lacking. Second, the best performance of such antireflection coating lags that of the conventional thin-film antireflection technology. Third, research has been conducted in application to a few specific materials, and the versatility of this approach remains unclear.

To address the aforementioned issues, we demonstrate the nanoscale optical coupling and photocurrent enhancement with dielectric nano-resonators atop various solar cell materials (e.g., Si and GaAs). To examine nanoscale optoelectronic responses, we employ a near-field scanning photocurrent microscopy technique. By utilizing an atomic force microscopy (AFM) probe with an attached optical fiber, we measure local photo-response of solar cells [11]-[13]. While the probe makes a raster scanning, topography and photocurrent of samples are obtained at the same time. Correlating with other characterization techniques, we find that the combination of a thin-film interference effect and resonant coupling of whispering gallery modes improves solar cells by more than 20 %. We also suggest novel hybrid nano-resonator arrays composed of various materials (e.g., SiO_2, silicon nitride (Si_3N_4), and titanium dioxide (TiO_2)) that can surpass the conventional thin-film-based antireflection technologies.

*dongheon.ha@nist.gov; phone 1-301-975-4557; www.nist.gov/cnst

Ha, Dongheon; Zhitenev, Nikolai.
"Improving dielectric nano-resonator-based antireflection coatings for photovoltaics."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

2. RESULTS AND DISCUSSION

We deposit close-packed 700 nm SiO_2 nano-resonators as antireflection coatings atop Si and GaAs solar cells to improve their optical and electrical properties. Nano-resonators are deposited by a Meyer-rod rolling technique [11], [12], [14]. By adjusting either a concentration of suspension containing SiO_2 nano-resonantors or the size of a Meyer-rod, a monolayer coating can be made. Deposited nano-resonators have a finite variation of diameters: 697 nm ± 80 nm. We measure light absorptivity enhancement with such nano-resonators (see Fig. 1a and d). Absorptivity enhancement is determined by comparing light absorption with nano-resonators with light absorption of bare samples. We also determine absorptivity enhancement from finite-difference time-domain (FDTD) calculations (see Fig. 1b and e).

Enhanced light absorption is observed with 700 nm SiO_2 nano-resonators in macroscopic measurements and FDTD calculations. While mostly broadband features are visible in the measurements, both broadband and narrowband features are clearly shown in the calculations. The broadband features can be explained by the thin-film interference effect. We model a thin-film layer atop Si and GaAs substrate and determine how this layer affects optical responses (see Fig. 1c and f). The thin-film layer has an effective refractive index varying in vertical position to mimic the properties of the close-packed SiO_2 nano-resonators: the effective refractive index of the film is calculated by averaging refractive indices of air and SiO_2 according to the geometric fraction of two materials in close-packed configurations [11], [12]. The thin-film model effectively describes broadband absorptivity enhancement observed in the experiments and calculations.

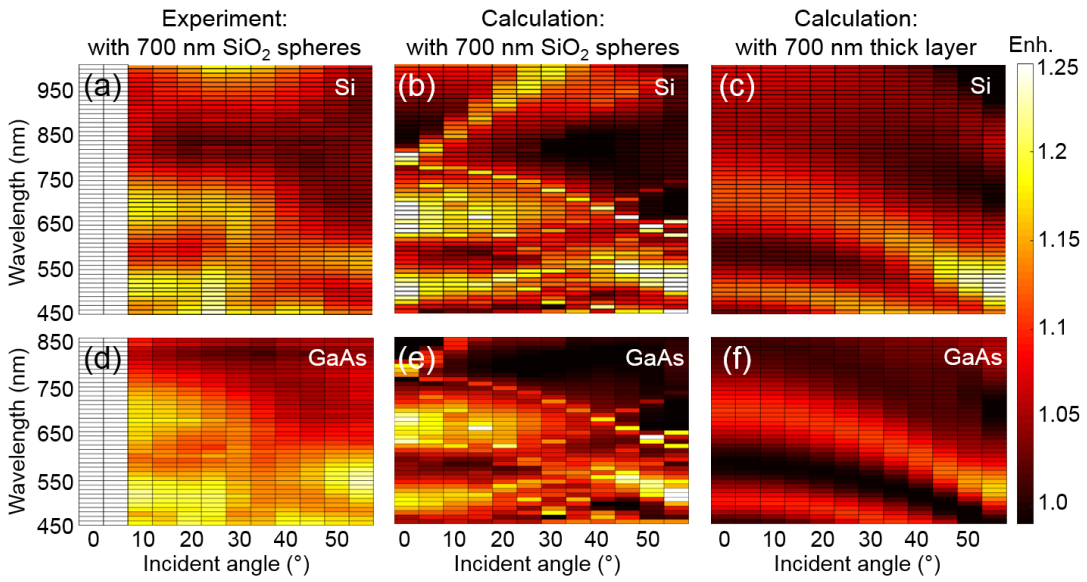


Figure 1. Absorptivity enhancements determined by macroscopic optical measurements for (a) a Si solar cell and (d) a GaAs solar cell with deposited SiO_2 nano-resonators with a mean diameter of 697 nm and a standard deviation of 80 nm. Absorptivity enhancements determined by FDTD calculations for (b) a Si solar cell with close-packed 700 nm SiO_2 nano-resonators, (c) a Si solar cell with a 700 nm thick thin-film layer, (e) a GaAs solar cell with close-packed 700 nm SiO_2 nano-resonators, and (d) a GaAs solar cell with a 700 nm thick thin-film layer.

Narrowband features can be explained by the excited whispering gallery modes within nano-resonators. We calculate the electric field intensity profiles at wavelengths where significant absorptivity enhancement is observed (see Fig. 2). Electric field profiles at wavelengths of 635 nm and 798 nm (marked with number 3 and 4) under transverse magnetic (TM) incident polarization show strong confined optical modes within nano-resonators. They are the excited whispering gallery modes responsible for the narrowband absorptivity enhancement at the corresponding wavelengths. However, such enhancement is barely seen in the macroscopic measurement due to the size variation of the nano-resonators [11], [12]. Repeated readings under identical conditions yields maximum measurement variation less than 1 % [11], [12].

Excited whispering gallery modes can be tuned by the size of nano-resonators (see Fig. 2). They are either blue- or red-shifted as the size of nano-resonators become smaller or larger: the excited whispering gallery modes at the wavelengths

Ha, Dongheon; Zhitenev, Nikolai.
"Improving dielectric nano-resonator-based antireflection coatings for photovoltaics."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

of 635 nm with 700 nm nano-resonators (marked with number 3) appear at the wavelength of 455 nm for 500 nm nano-resonators (marked with number 1) and at the wavelength of 906 nm for 1000 nm nano-resonators (marked with number 5), and the modes at the wavelength of 798 nm with 700 nm nano-resonators (marked with number 4) is reproduced at the wavelength of 573 nm for 500 nm nano-resonators (marked with number 2).

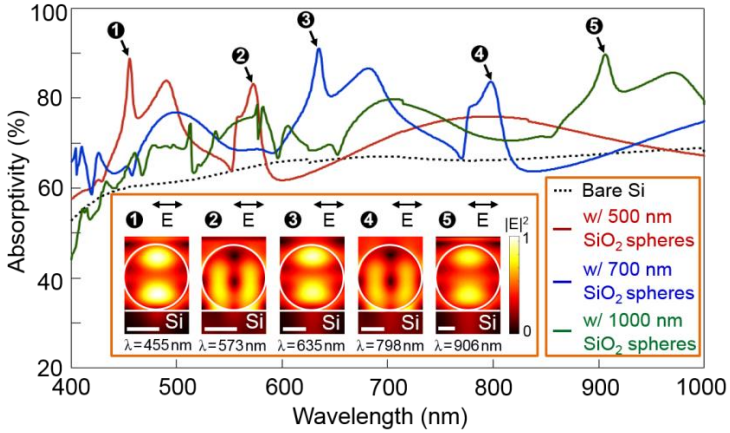


Figure 2. Absorptivity profile of a Si solar cell with close-packed 500 nm (red), 700 nm (blue), and 1000 nm (green) nano-resonators. Inset: electric field profile intensity profiles show excited whispering gallery modes at corresponding wavelengths under TM incident polarization (see matched number in absorptivity profiles). Scale bars are 250 nm.

Despite the marked enhancement with such nano-resonators, however, there is no clear experimental evidence confirming the narrowband features that originate from the whispering gallery modes. Only broadband features are seen in macroscale measurements. Therefore, we perform nanoscale photocurrent measurements using near-field photocurrent microscopy (NSPM) to experimentally demonstrate the photo-response enhancement induced by the excitation of whispering gallery modes.

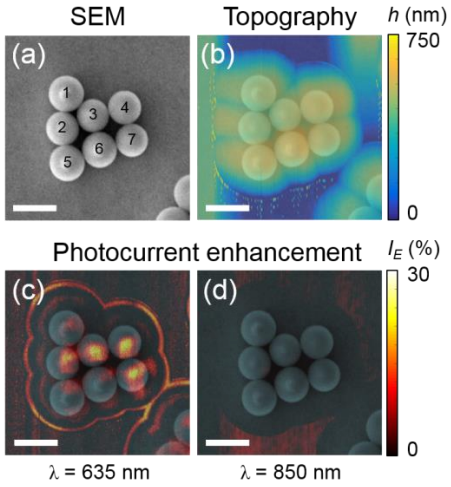


Figure 3. (a) An SEM image showing an area selected for NSPM measurements. Nano-resonators with numbers 2, 3, 4, and 7 have their size of ≈ 700 nm, and nano-resonators with numbers 1, 5, and 6 have their size of ≈ 750 nm. Both (b) topography and (c, d) photocurrent images are simultaneously obtained during the NSPM measurements. Photocurrent enhancement (c) at a wavelength for the excitation of whispering gallery modes and (d) at an off-resonance wavelength. Scale bars are 1 μm.

During the NSPM measurements, both topography and photocurrent of the sample are simultaneously obtained from the selected area of a GaAs solar cell with seven nano-resonators (see Fig. 3). Photocurrent enhancement is determined by normalizing the measured photocurrent to the photocurrent obtained in a bare area. NSPM measurements are performed at a resonant wavelength (λ = 635 nm) and an off-resonance wavelength (λ = 850 nm) that are found from the FDTD calculations (see Fig. 1b, e, and Fig. 2). A value of photocurrent enhanced by more than 30 % is measured over nano-resonators with the diameter ≈ 700 nm at the wavelength of the excitation of whispering gallery modes (λ = 635 nm), and the enhancement is smaller as the size deviates from 700 nm. At an off-resonance wavelength (λ = 850 nm), no specific photocurrent enhancement is detected with such nano-resonators.

As the optical resonances depend on the size of nano-resonators (Fig. 2), it is necessary to find an optimal size that can offer the maximum light absorption averaged across the solar spectrum. We first determine how different diameters affect the absorptivity peaks on a Si solar cell (see Fig. 4a). As previously discussed, in larger nano-resonators the resonance peaks shift to the longer wavelengths. For nano-resonators with diameters larger than ≈ 700 nm, second order resonance peaks are observed at shorter wavelengths. Next, we determine the current densities weighted with air mass 1.5 global (AM1.5G) solar spectrum. The second order peaks contribution results in the higher photocurrent densities as the diameter increases (see Fig. 4c). Based on Fig. 4c, we conclude that the optimal diameter of SiO_2 nano-resonators for a Si solar cell is ≈ 940 nm, where we calculate the highest AM1.5G solar spectrum-weighted photocurrent density, ≈ 31 mA/cm^2.

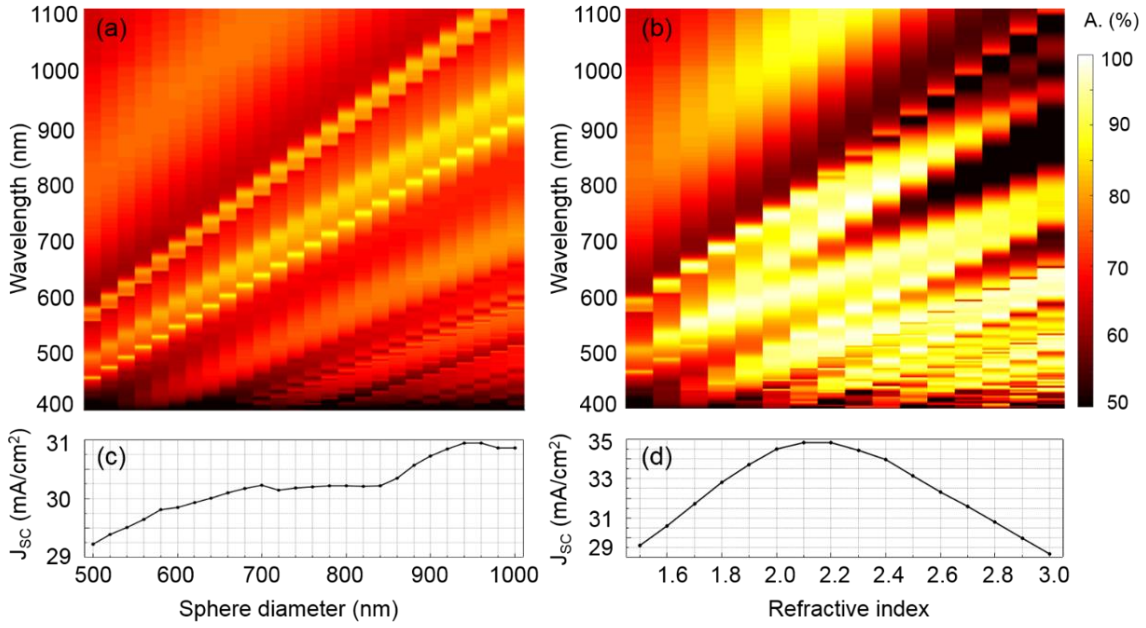


Figure 4. Absorptivity with nano-resonators can vary depending on (a) the size or (b) the refractive index of materials. Corresponding solar spectrum-weighted current densities also vary depending on (c) the size or (d) the refractive index of materials.

Alternatively, the optical resonances can be tuned by the refractive index of materials for nano-resonators [12]. To find the optimal refractive index, we fix the size of nano-resonators to 500 nm. Then, the optimal refractive index corresponding to the highest AM1.5G solar spectrum-weighted photocurrent densities is ≈ 2.1. This value is about the same as the refractive index of Si_3N_4, the optimal material for a single layer thin-film antireflection coating at the air-Si interface.

Despite the marked photo-response enhancement with nano-resonators, the maximum achievable enhancement still lags behind thin-film-based antireflection technologies. To further improve optical responses, we suggest novel hybrid arrays of nano-resonators that are composed of three different materials. We combine nano-resonators made of SiO_2, Si_3N_4, and TiO_2 to excite more photonic resonance modes within the arrays [12]. This results in the photo-response enhanced by

more than 30 % compared to a bare cell, surpassing the maximum achievable enhancement with a single layer thin-film antireflection coating (e.g., ≈ 80 nm thick Si_3N_4 thin-film atop a Si solar cell). Table 1 summarizes the enhancement values expected from each type of nano-resonators-based antireflection coating atop a Si solar cell.

Table 1. Solar spectrum-weighted current density and its enhancement from a bare Si solar cell for each type of nano-resonators-based antireflection coating

Antireflection coating type on a Si solar cell	**Solar spectrum-weighted current density (mA/cm^2)**	**Enhancement from a bare Si solar cell (%)**
Bare Si (without an antireflection coating)	27.7	-
500 nm SiO_2 nano-resonators	29.2	5.4
700 nm SiO_2 nano-resonators	30.2	9.0
1000 nm SiO_2 nano-resonators	30.9	11.6
500 nm Si_3N_4 nano-resonators	34.9	26.0
500 nm TiO_2 nano-resonators	29.5	6.5
Mixed combination of 500 nm SiO_2, TiO_2, and Si_3N_4 nano-resonators (equally distributed)	36.2	30.7

3. CONCLUSIONS

Various dielectric nano-resonators-based antireflection coatings have been introduced. With SiO_2 nano-resonator arrays placed atop a Si substrate, significant absorptivity enhancement is observed due to the combined effect of the excited whispering gallery modes and the thin-film interference. The spectrum of the whispering gallery modes is tunable by the diameter of SiO_2 nano-resonators. The effect of refractive index of nano-resonators on optoelectronic properties is also investigated. Finally, novel configurations of hybrid arrays made of SiO_2, Si_3N_4, and TiO_2 nano-resonators are introduced. The arrays show substantial absorptivity and photocurrent enhancements due to larger number of resonant and photonic modes. The use of dielectric nano-resonator arrays as antireflection coatings can reduce the cost and/or improve efficiency of solar cells, as these coatings do not require either surface patterning or complicated fabrication processes.

ACKNOWLEDGEMENTS

D. Ha acknowledges support under the Cooperative Research Agreement between the University of Maryland and the Center for Nanoscale Science and Technology at the National Institute of Standards and Technology, Award 70NANB14H209, through the University of Maryland.

REFERENCES

[1] Yan, R., Gargas, D., Yang, P., "Nanowire photonics," Nat. Photonics 3, 569-576 (2009).
[2] Atwater, H. A. and Polman, A., "Plasmonics for improved photovoltaic devices," Nat. Mater. 9, 205-213 (2010).
[3] Priolo, F., Gregorkiewicz, T., Galli, M., and Krauss, T. F., "Silicon nanostructures for photonics and photovoltaics," Nat. Nanotechnol. 9, 19-32 (2014).
[4] Polman, A. and Atwater, H. A., "Photonic design principles for ultrahigh-efficiency photovoltaics," Nat. Mater. 11, 174-177 (2012).

[5] Spinelli, P., Ferry, V. E., van de Groep, J., van Lare, M., Verschuuren, M. A., Schropp, R. E. I., Atwater, H. A., and Polman, A., "Plasmonic light trapping in thin-film Si solar cells," J. Opt. 14, 024002 (2012).
[6] Brongersma, M. L., Cui, Y., and Fan, S., "Light management for photovoltaics using high-index nanostructures," Nat. Mater. 13, 451-460 (2014).
[7] Ha, D., Fang, Z., Hu, L., and Munday, J. N., "Paper-based anti-reflection coatings for photovoltaics," Adv. Energy Mater., 4, 1301804 (2014).
[8] Fang, Z., Zhu, H., Yuan, Y., Ha, D., Zhu, S., Preston, C., Chen, Q., Li, Y., Han, X., Lee, S., Chen, G., Li, T., Munday, J., Huang, J., and Hu, L., "Novel nanostructured paper with ultrahigh transparency and ultrahigh haze for solar cells," Nano Lett., 14, 765-773 (2014).
[9] Ha, D., Murray, J., Fang, Z., Hu, L., and Munday, J. N., "Advanced broadband antireflection coatings based on cellulose microfiber paper," IEEE J. Photovolt., 5, 577-583 (2015).
[10] Ha, D., Fang, Z., and Zhitenev, N. B., "Paper in electronic and optoelectronic devices," Adv. Electron. Mater., 4, 1700593 (2018).
[11] Ha, D., Yoon, Y., and Zhitenev, N. B., "Nanoscale imaging of photocurrent enhancement by resonator array photovoltaic coatings," Nanotechnology, 29, 145401 (2018).
[12] Ha, D. and Zhitenev, N. B., "Improving dielectric nano-resonator array coatings for solar cells," Part. Part. Syst. Charact. 35, 1800131 (2018).
[13] Yoon, Y., Ha, D., Park, I. J., Haney, P. M., Lee, S., and Zhitenev, N. B., "Nanoscale photocurrent mapping in perovskite solar cells," Nano Energy, 48, 543-550 (2018).
[14] Ha, D., Gong, C., Leite, M. S., and Munday, J. N., "Demonstration of resonance coupling in scalable dielectric microresonator coatings for photovoltaics," ACS Appl. Mater. Interfaces, 8, 24536-24542 (2016).

Ha, Dongheon; Zhitenev, Nikolai.
"Improving dielectric nano-resonator-based antireflection coatings for photovoltaics."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

Nanoimaging of local photocurrent in hybrid perovskite solar cells via near-field scanning photocurrent microscopy

Dongheon Ha*[a,b], Yohan Yoon[a,b], Ik Jae Park[c], Paul M. Haney[a], Nikolai B. Zhitenev[a]
[a]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD, USA 20899; [b]Maryland Nanocenter, University of Maryland, College Park, MD, USA 20742; [c]Department of Materials Science and Engineering, Seoul National University, Seoul, Korea 08826

ABSTRACT

Photocurrent generation of methylammonium lead iodide ($CH_3NH_3PbI_3$) hybrid perovskite solar cells is observed at the nanoscale using near-field scanning photocurrent microscopy (NSPM). We examine how the spatial map of photocurrent at individual grains or grain boundaries is affected either by sample post-annealing temperature or by extended light illumination. For NSPM measurements, we use a tapered fiber with an output opening of 200 nm in the Cr/Au cladded metal coating attached to a tuning fork-based atomic force microscopy (AFM) probe. Increased photocurrent is observed at grain boundaries of perovskite solar cells annealed at moderate temperature (100 °C); however, the opposite spatial pattern (i.e., increased photocurrent generation at grain interiors) is observed in samples annealed at higher temperature (130 °C). Combining NSPM results with other macro-/microscale characterization techniques including electron microscopy, x-ray diffraction, and other electrical property measurements, we suggest that such spatial patterns are caused by material inhomogeneity, dynamics of lead iodide segregation, and defect passivation. Finally, we discuss the degradation mechanism of perovskite solar cells under extended light illumination, which is related to further segregation of lead iodide.

Keywords: Perovskite, grain boundaries passivation, near-field scanning optical microscopy (NSOM), near-field scanning photocurrent microscopy (NSPM), nanoscale measurement, lead iodide, light-induced degradation, degradation mechanism

1. INTRODUCTION

Finding ways to make cost-effective photovoltaics (PVs) is one of the most significant tasks to address current societal issues, such as the greenhouse effect, air pollution, exhaustion of traditional energy sources, etc. Tandem photovoltaics have been intensively investigated in order to achieve high power conversion efficiency [1], [2]. However, such cells are still costly and limited by materials selection. Also, they require complicated fabrication processes preventing their commercialization as cost-effective energy sources. Nanophotonic engineering has also been extensively studied to increase cost-effectiveness of photovoltaics. High power conversion efficiencies have been observed with low-cost, environmentally-friendly, and earth-abundant nano-materials (e.g., micro-/nanosize cellulose fibers, nanosize glass beads, etc.), with reported improvement of more than 20 % over bare cells [3]-[6]. However, more studies of durability (both mechanical and environmental) are still needed to fully commercialize such approaches [7].

Perovskite solar cells have attracted significant interest in the PV research community due to their easy fabrication process, low-cost, and the surprisingly fast progress of power conversion efficiency. Currently, the efficiency of the best perovskite solar cell is comparable to that of traditional multi-crystalline semiconductor-based PVs (e.g., cadmium telluride, silicon, etc.) [8], [9]. However, perovskite solar cells also have multiple instability (e.g., degradation, hysteresis, etc.) issues driven by several environmental factors, such as light illumination, humidity, temperature, etc. Such issues have impeded the broad deployment of perovskite solar cells as commercial PVs. Therefore, the origin of the instability should be further examined. Other unknown factors contributing to the degradation should be identified, and potentially hazardous side-effect should also be addressed as this technology involves toxic materials (e.g., lead).

Multiple characterization techniques have been applied to understand in-depth compositional/structural properties and

*dongheon.ha@nist.gov; phone 1 301 975-4557; www.nist.gov/cnst

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoimaging of local photocurrent in hybrid perovskite solar cells via near-field scanning photocurrent microscopy."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

operational principles of perovskite solar cells [10]–[12]. However, to fully understand the operational principles and/or the causes of the aforementioned instability, detailed studies should be made at the nanoscale, where photo-excited carrier generation and collection actually take place.

In this work, we investigate photocurrent generation and collection of perovskite solar cells at the nanoscale using a near-field scanning photocurrent microscopy (NSPM) technique. We observe opposite carrier collection spatial patterns related to the post-annealing temperature of samples. The nanoscale measurement is correlated with other macro-/micro characterization techniques, such as x-ray diffraction, electron microscopy, current density-voltage measurement, etc. Based on the combined characterizations, we attribute the different nanoscale behavior to the dynamics of lead iodide (PbI_2) segregation and defect passivation. Finally, we also discuss how an extended light illumination causing the cell degradation affects the microstructure and the local carrier collection.

2. RESULTS AND DISCUSSION

We fabricate perovskite solar cells and characterize their properties at the macro- and nanoscale. We employ a two-step spin-coating method to deposit methylammonium lead iodide perovskite layer ($CH_3NH_3PbI_3$, $MAPbI_3$ hereafter) on a hole transport layer made from nickel oxide (NiO_x). The hole transport layer is coated on an indium tin oxide (ITO) substrate. An electron transport layer is made from [6,6]-phenyl C61 butyric acid methyl ester (PCBM), and a 120 nm thick silver (Ag) contact is formed atop the PCBM layer for carrier extraction. The cross-sectional scanning electron microscopy (SEM) image of the fabricated sample is shown in Fig. 1a.

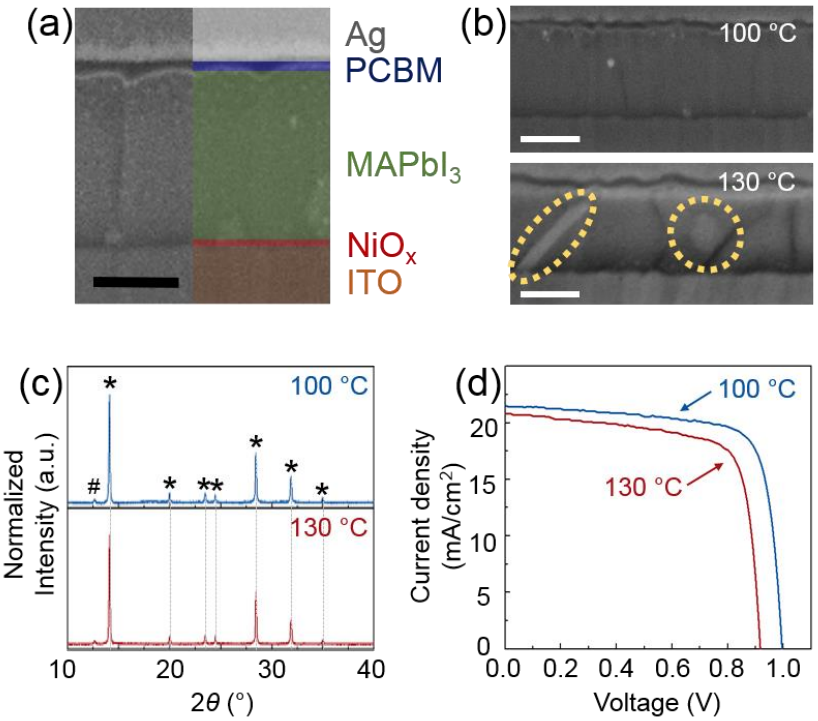


Figure 1. (a) A cross-sectional SEM image showing the structure of the fabricated perovskite solar cell. Scale bar is 250 nm. (b) Cross-sectional SEM images showing the perovskite layers annealed at 100 °C (top) and 130 °C (bottom), respectively. Scale bars are 250 nm. (c) X-ray diffraction patterns (#: PbI_2, *: $MAPbI_3$) and (d) current density-voltage curves for samples annealed at 100 °C (blue) and 130 °C (red), respectively.

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoimaging of local photocurrent in hybrid perovskite solar cells via near-field scanning photocurrent microscopy."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

We applied two different post-annealing temperatures, 100 °C and 130 °C, to form crystalline structures. Depending on the annealing temperature, we observe different microstructure of lead iodide (PbI_2) within the samples (see SEM images in Fig. 1b). In the sample annealed at higher temperature (130 °C), we find distinct and bright spots at grain boundaries (see spots marked with yellow dotted circles). They can be identified as PbI_2 crystallites due to their low conductivity and the accumulation of electrons during SEM imaging processes. However, the distribution of PbI_2 is relatively uniform across the sample that is annealed at moderate temperature (100 °C) and no significant PbI_2 crystallites are found.

Despite the different distribution of PbI_2 crystallites, both samples have similar intensities of PbI_2 lines in x-ray diffraction patterns (see Fig. 1c). The intensities of perovskite (i.e., $MAPbI_3$) lines are also similar. Current density-voltage (*J-V*) characteristics show both higher current density and voltage in the sample annealed at 100 °C (see Fig. 1d). This corresponds to 14.65 % higher power conversion efficiency for this sample (16.98 % vs. 14.81 %). Repeated measurements under identical conditions yield maximum variation of less than 1 % [13].

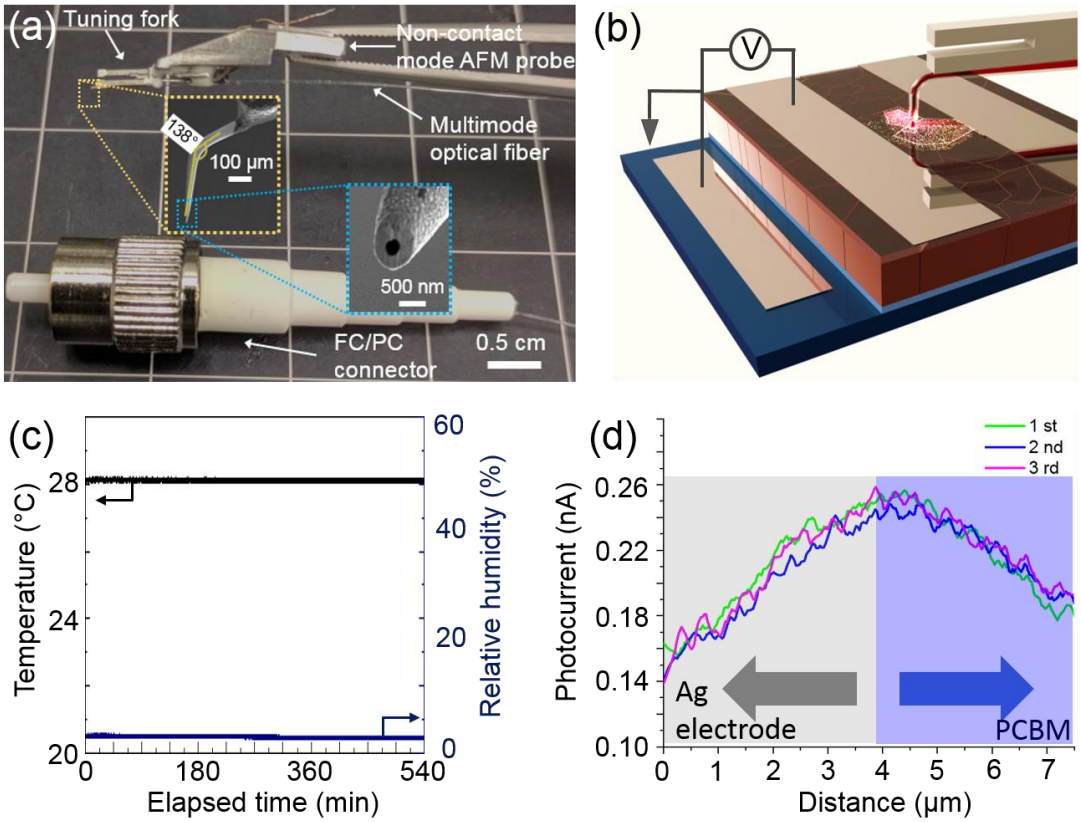


Figure 2. (a) A picture showing the NSPM probe. (b) A schematic describing the NSPM measurement technique. (c) Time dependent environmental factors (temperature and relative humidity) inside the AFM chamber when NSPM measurements are performed. (d) Line scan profiles of NSPM measurements near the Ag electrode of the sample.

To shed light on the role of the post-annealing, we investigate photovoltaic operation at the nanoscale where photo-excited carrier generation and collection take place. We employ a near-field scanning photocurrent microscopy technique to image photocurrent. In this measurement, a tuning fork-based non-contact mode atomic force microscopy (AFM) probe is used. A multimode optical fiber is attached to the probe to locally illuminate the samples [13]–[15]. The end of the optical fiber is coated with metal (20 nm of Cr and 200 nm of Au) and has ≈ 200 nm output hole for photon injection (see Fig. 2a). The other end of the optical fiber is connected to a bench top diode laser (a wavelength of 635 nm

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoimaging of local photocurrent in hybrid perovskite solar cells via near-field scanning photocurrent microscopy."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

or 780 nm) via a FC/PC connector. The NSPM probe has a spring constant of 15 N/m at a resonance frequency of ≈ 35 kHz.

For NSPM measurements, perovskite solar cells are placed onto a piezo stage within an AFM system. As the NSPM probe makes a raster scanning, topography and photocurrent signals are obtained at the same time (see Fig. 2b). To amplify the detected signals, a variable-gain low-noise amplifier is used with a gain of 10^7 A/V, and the amplified photocurrent is then detected by a lock-in technique. We keep the distance between the NSPM probe and the top surface of the perovskite solar cell at ≈ 10 nm during the measurements.

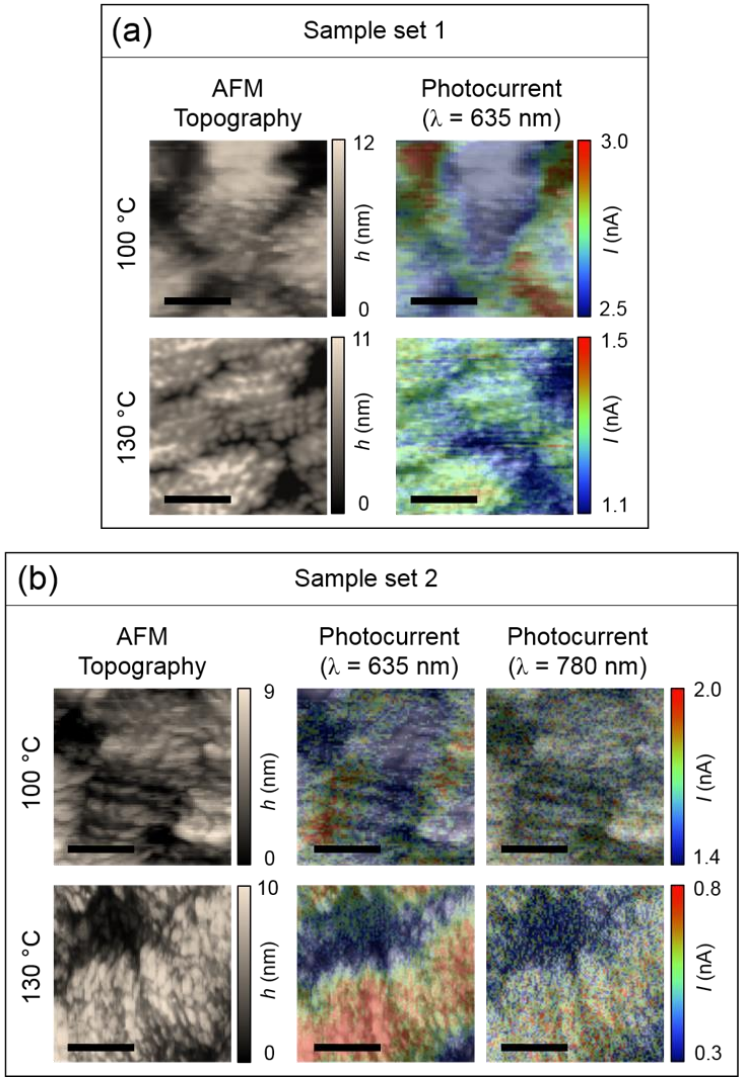


Figure 3. (a) NSPM images on the first set of samples annealed at 100 °C and 130 °C. Measurements are performed at a wavelength of 635 nm. (b) NSPM images on the second set of samples annealed at 100 °C and 130 °C. Measurements are performed at two different wavelengths: 635 nm and 780 nm. Scale bars are 500 nm.

During all measurements, environmental factors within the AFM chamber that can affect the cell operation (e.g., relative humidity and temperature) are accurately controlled (see Fig. 2c). NSPM measurements are performed near Ag contacts to increase signal to noise ratio (SNR). To find measurement areas that offer maximum SNRs, we first perform line

scans near the boundary between the Ag contacts and the PCBM layer. NSPM measurements are then carried out at the areas of maximal signal (see Fig. 2d). While performing NSPM measurements, laser power is optimized to maintain high enough SNR while avoiding sample degradation. Three repeated line scan profiles in Fig. 2d confirm no undesired light-induced sample degradation during the NSPM measurements.

A clear spatial pattern contrast is observed in NSPM images of samples annealed at two different temperatures (see results for first set of samples in Fig. 3a). Measurements are performed at a wavelength of 635 nm where the sun spectral irradiance is near maximum. In the sample annealed at 100 °C, higher photocurrent is generated and collected at grain boundaries. Meanwhile, the opposite spatial pattern is observed in the sample annealed at 130 °C: higher photocurrent at grain interiors and lower photocurrent at grain boundaries. Note that the photocurrent contrast is not caused by the thickness variation. Significant variation of photocurrent (more than 20 %) is observed with a limited height variation (≈ 10 nm) corresponding to less than 2.5 % variation in the absorber thickness. The increased photocurrent collection at grain boundaries of the sample annealed at 100 °C can be explained by the beneficial role of PbI_2 passivation [13], [16]. As the annealing temperature increases to 130 °C, however, more structural and functional changes from $MAPbI_3$ to PbI_2 takes place, forming distinguishable PbI_2 crystallites at grain boundaries, as confirmed by the SEM image in Fig. 1b. The segregated PbI_2 phase leads to an increased carrier recombination at grain boundaries, and therefore grain boundaries become less efficient carrier collectors for this sample. NSPM measurements of another set of samples at two different wavelengths: 635 nm and 780 nm (see Fig. 3b) confirm such behavior. Measurements at a wavelength 780 nm are additionally performed to observe the nanoscale photo-excited carrier generation and collection near the bandgap of the active material (i.e., perovskite) with light penetrating deeper in the absorber.

We also conduct NSPM measurements to understand the mechanism behind the instability and degradation of perovskite solar cells under normal light illumination. NSPM measurements are conducted in the multiple stages of the light-induced degradation process under extended light illumination (AM1.5G 0.1 sun, ≈ 25 % relative humidity in air at room temperature). We find that the degradation patterns under continuous light illumination proceed in different ways in two different samples. In the sample annealed at 100 °C, we observe suppressed carrier collection at some grain boundaries. More structural and compositional transformation from $MAPbI_3$ to PbI_2 occurs under the continuous light illumination, and the positive role of lead iodide passivation at such grain boundaries fades away. This seriously affects the overall performance of this sample, leading to rapid degradation of the photovoltaic properties under an extended light illumination. However, grain interiors remain the dominant carrier collection paths for the sample annealed at 130 °C regardless of the light illumination time. This results in a more robust behavior of this sample under an extended light illumination. In summary, the sample annealed at 130 °C is more robust under extended light illumination, and such behavior is closely related to the segregation of PbI_2 crystallites at grain boundaries.

3. CONCLUSIONS

We investigate nanoscale photo-excited carrier generation and collection of methylammonium lead iodide perovskite solar cells. Combining the results from nanoscale NSPM photocurrent measurements with other characterization techniques, we reveal the carrier dynamics, the role of lead iodide, the effect of sample preparation temperatures at the macro-/nanoscale. Enhanced photo-excited carrier collection is observed at grain boundaries in the sample annealed at 100 °C, while significantly higher photocurrent is measured at grain interiors in the sample annealed at 130 °C. Such spatial patterns are related to the material inhomogeneity and the dynamics of lead iodide segregation. We also suggest the degradation mechanism of perovskite solar cells under extended light illumination based on the results from NSPM measurements. The structural and compositional changes in perovskite layer under extended light illumination suppress the beneficial role of grain boundaries in the sample annealed at 100 °C.

ACKNOWLEDGEMENTS

D. Ha and Y. Yoon acknowledge support under the Cooperative Research Agreement between the University of Maryland and the Center for Nanoscale Science and Technology at the National Institute of Standards and Technology, Award 70NANB14H209, through the University of Maryland.

REFERENCES

[1] Kinga, R. R., Law, D. C., Edmondson, K. M., Fetzer, C. M., Kinsey, G. S., Yoon, H., Sherif, R. A., and Karam, N. H., "40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells," Appl. Phys. Lett., 90, 183516 (2007).

[2] Guter, W., Schöne, J., Philipps, S. P., Steiner, M., Siefer, G., Wekkeli, A., Welser, E., Oliva, E., Bett, A. W., and Dimroth, F., "Current-matched triple-junction solar cell reaching 41.1% conversion efficiency under concentrated sunlight," Appl. Phys. Lett. 94, 223504 (2009).

[3] Ha, D., Fang, Z., Hu, L., and Munday, J. N., "Paper-based anti-reflection coatings for photovoltaics," Adv. Energy Mater., 4, 1301804 (2014).

[4] Fang, Z., Zhu, H., Yuan, Y., Ha, D., Zhu, S., Preston, C., Chen, Q., Li, Y., Han, X., Lee, S., Chen, G., Li, T., Munday, J., Huang, J., and Hu, L., "Novel nanostructured paper with ultrahigh transparency and ultrahigh haze for solar cells," Nano Lett., 14, 765-773 (2014).

[5] Ha, D., Murray, J., Fang, Z., Hu, L., and Munday, J. N., "Advanced broadband antireflection coatings based on cellulose microfiber paper," IEEE J. Photovolt., 5, 577-583 (2015).

[6] Ha, D., Gong, C., Leite, M. S., and Munday, J. N., "Demonstration of resonance coupling in scalable dielectric microresonator coatings for photovoltaics," ACS Appl. Mater. Interfaces, 8, 24536-24542 (2016).

[7] Ha, D., Fang, Z., and Zhitenev, N. B., "Paper in electronic and optoelectronic devices," Adv. Electron. Mater., 4, 1700593 (2018).

[8] Green, M. A., Hishikawa, Y., Dunlop, E. D., Levi, D. H., Hohl-Ebinger, J., and Ho-Baillie, A. W. H., "Solar cell efficiency tables (version 51)," Prog. Photovolt. Res. Appl., vol. 26, 3-12 (2018).

[9] National Renewable Energy Laboratory (U.S. Department of Energy), Research-Cell Record Efficiency Chart, <https://www.nrel.gov/pv/assets/images/efficiency-chart-20180716.jpg> (2018).

[10] Yun, J. S., Ho-Baillie, A., Huang, S., Woo, S. H., Heo, Y., Seidel, J., Huang, F., Cheng, Y. -B., and Green, M. A., "Benefit of grain boundaries in organic–inorganic halide planar perovskite solar cells," J. Phys. Chem. Lett., 6, 875-880 (2015).

[11] Kim, J. H., Liang, P. -W., Williams, S. T., Cho, N., Chueh, C. -C., Glaz, M. S., Ginger, D. S., and Jen, A. K. -Y., "High-performance and environmentally stable planar heterojunction perovskite solar cells based on a solution-processed copper-doped nickel oxide hole-transporting layer," Adv. Mater., 27, 695-701 (2015).

[12] Bergmann, V. W., Weber, S. A. L., Ramos, F. J., Nazeeruddin, M. K., Grätzel, M., Li, D., Domanski, A. L., Lieberwirth, I., Ahmad, S., and Berger, R., "Real-space observation of unbalanced charge distribution inside a perovskite-sensitized solar cell," Nat. Commun., 5, 5001, 2014.

[13] Yoon, Y., Ha, D., Park, I. J., Haney, P. M., Lee, S., and Zhitenev, N. B., “Nanoscale photocurrent mapping in perovskite solar cells,” Nano Energy, 48, 543-550 (2018).

[14] Ha, D., Yoon, Y., and Zhitenev, N. B., “Nanoscale imaging of photocurrent enhancement by resonator array photovoltaic coatings,” Nanotechnology, 29, 145401 (2018).

[15] Ha, D. and Zhitenev, N. B., “Improving dielectric nano-resonator array coatings for solar cells,” Part. Part. Syst. Charact. 35, 1800131 (2018).

[16] Chen, Q., Zhou, H., Song, T. -B., Luo, S., Hong, Z., Duan, H. -S., Dou, L., Liu, Y., and Yang, Y., “Controllable self-induced passivation of hybrid lead iodide perovskites toward high performance solar cells,” Nano Lett., 14, 4158-4163 (2014).

Ha, Dongheon; Yoon, Yohan; Park, Ik Jae; Haney, Paul; Zhitenev, Nikolai.
"Nanoimaging of local photocurrent in hybrid perovskite solar cells via near-field scanning photocurrent microscopy."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing

Ran Tao[1], Aaron M. Forster[1], Kirk D. Rice[1], Randy A. Mrozek[2], Shawn T. Cole[2], Reygan M. Freeney[3]

[1]*Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA*
ran.tao@nist.gov
[2]*U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, USA*
[3]*U.S. Army Aberdeen Test Center, Aberdeen Proving Ground, MD, USA*

Abstract. The current standard backing material used for ballistic resistance testing of body armour, Roma Plastilina No. 1 (RP1), is known to exhibit complex thermomechanical behavior under actual usage conditions. Body armour test standards that specify RP1 often establish a performance requirement for the RP1 before it can be used. To meet this requirement, RP1 most likely must be temperature conditioned. This conditioning step adds to the complexity of the test and introduces temperature-time variables into the test, which serve as incentives for finding potential alternative materials that perform similarly at room temperature. To achieve the goal of replacing the current standard backing material, the U.S. Army Research Laboratory (ARL) and U.S. Army Aberdeen Test Center (ATC) have taken the lead to develop a family of room-temperature backing materials that exhibit dynamic properties that are consistent and similar to temperature-conditioned RP1. These candidate materials, named ARTIC, have fewer components than the RP1 clay. The research at NIST focuses on rheological characterization and thermal analysis of these next-generation backing materials to understand structure-property relationships and compare to the current RP1. In this work, we show that ARTIC materials exhibit minimal temperature dependence of mechanical properties and consistent thermal properties over a wide temperature range.

1. INTRODUCTION

Roma Plastilina No. 1 (RP1) is a manufactured oil-based clay that was chosen by both the National Institute of Justice (NIJ) and the U.S. Army as the standard backing material for ballistic resistance testing of body armour [1-3]. Body armour is tested by placing the armour against a backing material, also known as a ballistic witness material (BWM). After a ballistic impact, evaluation of the armour's performance is based on assessing perforation of the armour and deformation depth of the backing material. RP1 clay is primarily a sculpting clay that is a multicomponent, multiphase formulation composed of oils, waxes, and clay minerals with additives. Unfortunately, over the decades since RP1 was first adopted as the standard, changes have been made to the RP1 formulation by the clay manufacturer. Those changes were not driven by BWM performance requirements, but to meet the demands of other users, primarily from artist communities. As a result, newer versions of RP1 are stiffer at room temperature than the original RP1 selected by NIJ and the Army. In order to meet the specifications, ballistics researchers and practitioners must now thermally condition the material prior to use, which involves heating RP1 to approximately 40 °C or higher.

A dynamic test procedure to verify that clay can be used for ballistic testing of body armour is called a "drop test", which involves dropping an impactor of specified mass and geometry from a specified height onto the surface of the clay box and then measuring the indentation depth to ensure that it falls within specification limits. In addition to the undesired extra thermal conditioning step, the mechanical properties of clay depend on work and thermal history, temperature, and time [4-7]. The complex relationships between these factors have led to an initiative to develop an alternative backing material to replace RP1, as recommended in the National Research Council reports [8, 9]. The major requirements of the alternative backing material are that it provides desirable, predictable, and controllable properties. Our motivation for

Tao, Ran; Forster, Aaron; Rice, Kirk; Mrozek, Randy A.; Cole, Shawn; Freeney, Reygan.
"Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing."
Paper presented at Personal Armour Systems Symposium 2018, Washington, D.C., United States. October 1, 2018 - October 5, 2018.

this research is to understand structure-processing-property relationships for BWM candidate materials and compare to the current RP1.

The Army Research Laboratory (ARL) has taken the lead to develop a replacement backing material, named ARTIC, and initial success has been achieved [10-11]. The name ARTIC is taken from ARL reusable, temperature insensitive "clay" [10]. The ARTIC material has fewer components, which is expected to show more consistent behavior. Freeney and Mrozek [10] reported that formulation of ARTIC can be tuned to match the behavior of RP1 at 100 °F (37.8 °C) in force penetration experiments, and they show that ARTIC materials are room temperature materials with no temperature dependence up to 100 °F. Those ARTIC materials satisfied the drop test requirements and yielded ballistic test results that were statistically similar to those obtained with RP1 as the BWM [10-11]. Furthermore, ARTIC materials show no ageing effect; the ballistic and indentation response does not change with time. To help with evaluation of the ARTIC material, we use thermo-rheological characterization methods to investigate the backing material.

2. EXPERIMENTAL[1]

2.1 Materials

Roma Plastilina No. 1 (RP1) clay was used as received from the manufacturer (Sculpture House, Springhill, NJ). The ARTIC materials were provided by the Army Research Laboratory (Aberdeen Proving Ground, MD) and were designated as ARTIC 5.5, ARTIC 6.5, and ARTIC 8.0 based on different formulations.

2.2 Characterization Methods

2.2.1 Differential scanning calorimetry (DSC)

Calorimetric measurements were performed under nitrogen flow using a TA Q2000 differential scanning calorimeter (TA Instruments, New Castle, DE). Two methods were employed: Method 1) A one-step heating scan at 10 °C/min from 25 °C to 300 °C and Method 2) a two-step heating scan at 10 °C/min from 25 °C to 130 °C, followed by an isotherm for 30 min at 130 °C to ensure dehydration, then heating up to 400 °C at 10 °C/min. For Method 1, hermetic aluminum pans were used; for Method 2, aluminum pans with pin hole lids were used. A second heating scan was also performed after cooling from 300 °C (Method 1) or 400 °C (Method 2) to 25 °C at 10 °C/min.

2.2.2 Rheology

Rheological experiments were performed using a rubber process analyzer RPA *elite* (TA Instruments, New Castle, DE) with enhanced air cooling system. The first advantage of the rubber process analyzer for measuring BWMs is that is offers a higher torque range for solid materials than that of a commercial rheometer, so that information under large deformation can be obtained. Second, it provides a consistent sample loading procedure, as ensured by the pressure pneumatic system together with the automated gap closure, such that the loading effects from sample to sample are minimized.

[1] Certain commercial equipment, instruments, or materials are identified in this presentation in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Tao, Ran; Forster, Aaron; Rice, Kirk; Mrozek, Randy A.; Cole, Shawn; Freeney, Reygan.
"Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing."
Paper presented at Personal Armour Systems Symposium 2018, Washington, D.C., United States. October 1, 2018 - October 5, 2018.

Figure 1. Test specimen (Roma Plastilina No. 1) is placed between two sheets of polyester film for rheological tests using rubber process analyzer. The image shows the specimen in a pressed form after the test is done.

The specimen was cut in the form of a rectangular parallelepiped from a larger block of material as received using a stiff blade and weighs approximately 5 g to ensure consistency. The test specimen was then placed between two sheets of polyester film and loaded onto the lower platen of the instrument for measurements (Figure 1). Upon initiating the test, the platens automatically close to compress the test specimen between the biconical platens. Frequency sweep experiments were performed at 0.01° strain (equivalent to 0.14 % strain) from 0.05 Hz to 50 Hz at five temperatures: 20 °C, 25 °C, 30 °C, 40 °C, and 50 °C, which covers the operating temperature range for current RP1 calibration. Three samples were tested at 25 °C to obtain the standard deviation of the results.

3. RESULTS

Figure 2 shows the normalized DSC heat flow responses measured on heating at 10 °C/min for the three ARTIC materials. The dashed lines correspond to the results obtained from Method 1, one-step heating scan from 25 °C to 300 °C (see Section 2.2.1). First, there is no thermal transition in the temperature range of interest, i.e., 25 °C to 50 °C, which is in contrast to the complex thermal behavior of the RP1 clay [5]. Second, an endothermic melting peak is observed for all three ARTIC materials upon the first heating scan. The endothermic peak with a maximum at ≈ 290 °C relates to the cleavage of macromolecular chains, resulting in thermal degradation [12]. On the second heating scan, no melting transition is present (results not shown for brevity). The solid lines represent the results from a two-step heating method by first heating up to 130 °C and holding for 30 min, then heating up to 400 °C at 10 °C/min. The two-step DSC results are identical for all ARTIC materials, which is as expected because those ARTIC materials have essentially the same constituents, and the only difference is the amount of solids loading. No melting transition is observed before degradation, indicating that the melting peaks seen in Method 1 may be due to moisture involved structures [13]. The sample dehydrates during the isothermal holding step at 130 °C using a DSC pan with a pin hole lid.

Tao, Ran; Forster, Aaron; Rice, Kirk; Mrozek, Randy A.; Cole, Shawn; Freeney, Reygan.
"Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing."
Paper presented at Personal Armour Systems Symposium 2018, Washington, D.C., United States. October 1, 2018 - October 5, 2018.

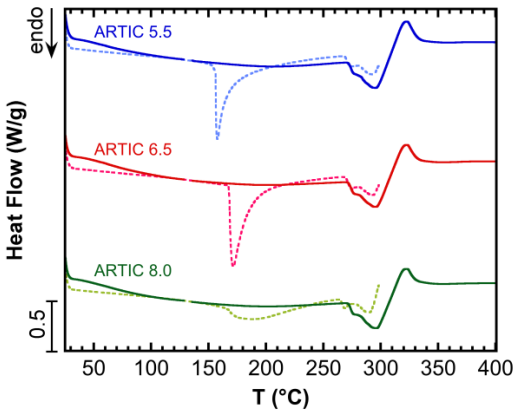


Figure 2. DSC heating scans at 10 °C/min for the ARTIC materials. The dashed and solid lines are results from Method 1 and Method 2, respectively. See Section 2.2.1 for details.

Figure 3 shows a double logarithmic plot of the dynamic storage modulus (G') as a function of frequency for the ARTIC materials and the RP1 clay. The trend of G' follows ARTIC 8.0 > ARTIC 6.5 > ARTIC 5.5, consistent with ARL's plateau modulus results from force penetration tests [10]. Second, the modulus of the RP1 clay shows a clear temperature dependence, i.e., G' decreases as temperature increases [7]. At 2 Hz, G' decreases from 14 MPa at 20 °C to 3 MPa at 50 °C for RP1 (Figure 4). For the ARTIC materials, the modulus data at different temperatures almost overlap with each other for each sample, indicating that the ARTIC materials are temperature insensitive within the measured temperature range (20 °C to 50 °C). In other words, ARTIC materials could be considered "room-temperature" backing materials, as shown in Figure 4 where G' at 2 Hz exhibits minimal temperature dependence. Furthermore, a stronger frequency dependence is observed for the RP1 clay. A weaker frequency dependency, as reflected in the ARTIC results, may be beneficial to reduce variations in material response when tested under different conditions.

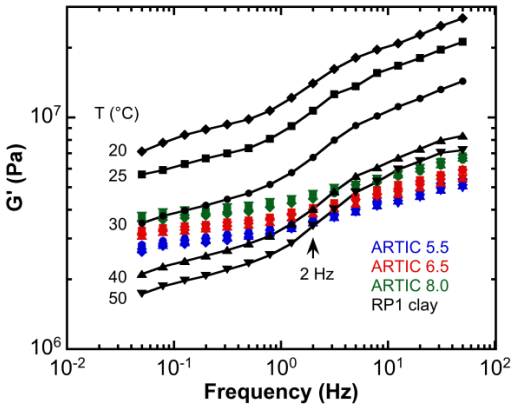


Figure 3. Dynamic shear storage modulus (G') as a function of frequency measured at a strain of 0.14 % (within linear range) at different temperatures ranging from 20 °C to 50 °C for the ARTIC materials and the RP1 clay [7]. Different symbols represent data points at different temperatures.

The RP1 clay is typically conditioned at or above 100 °F (~37.8 °C) to satisfy the drop test verification requirement. The G' data at different temperatures for all three ARTIC materials fall within the range of the 40 °C data of the RP1 over the same frequency range, suggesting that the ARTIC materials show promising properties as alternative BWMs. This also suggests that our approach of measuring G' may be

Tao, Ran; Forster, Aaron; Rice, Kirk; Mrozek, Randy A.; Cole, Shawn; Freeney, Reygan.
"Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing."
Paper presented at Personal Armour Systems Symposium 2018, Washington, D.C., United States. October 1, 2018 - October 5, 2018.

used as a validation method to narrow down the formulations of ARTIC, similar to ARL's force penetration experiments where the responses of ARTIC were found to match the results of the RP1 clay at 40 °C [10-11].

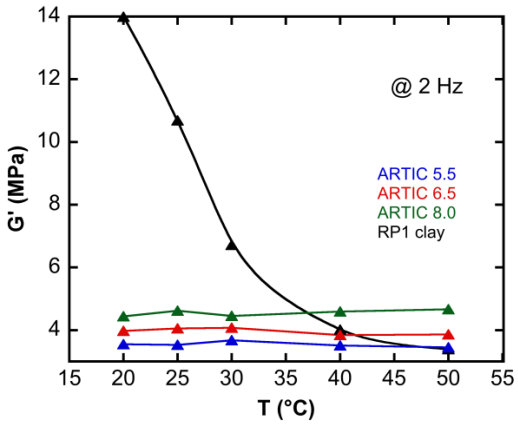


Figure 4. Dynamic shear storage modulus (G') as a function of temperature at 2 Hz from frequency sweep experiments (data extracted from Figure 3).

4. CONCLUSIONS

In this work, we used thermo-rheological methods to characterize the current standard backing material for body armour testing, Roma Plastilina No. 1 (RP1) clay, and a family of candidate backing materials, designated as ARTIC, developed by the Army Research Laboratory. The thermal studies show that there is no thermal transition in the temperature range of interest for the ARTIC materials, i.e., 20 °C to 50 °C, in contrast to the RP1 clay that exhibits complex thermal behaviour due to its multiphase formulation. The melting peaks observed using a hermetic pan for the ARTIC materials are absent after dehydrating the materials, indicating that the melting transitions may be caused by moisture involved structures. The effects of temperature on the rheological properties were investigated. The results show that ARTIC materials are promising room-temperature ballistic witness materials, as shown by minimal temperature dependence of the shear modulus and the agreement of the modulus data with those of the RP1 clay at the conditioning temperature for validation tests.

Acknowledgments

The authors thank Mr. Anicet Djakeu Samen for help with the rheological measurements.

REFERENCES

[1] R.N. Prather, C.L. Swann, C.E. Hawkins, Backface Signatures of Soft Body Armors and the Associated Trauma Effects, Chemical Systems Laboratory, Aberdeen Proving Ground, MD, 1977.

[2] US Department of Justice, Office of Justice Programs, National Institute of Justice, Ballistic Resistance of Body Armor NIJ Standard-0101.06, 2008.

Tao, Ran; Forster, Aaron; Rice, Kirk; Mrozek, Randy A.; Cole, Shawn; Freeney, Reygan.
"Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing."
Paper presented at Personal Armour Systems Symposium 2018, Washington, D.C., United States. October 1, 2018 - October 5, 2018.

[3] DoD Testing Requirements for Body Armor, Report Number D-2009-047. US Department of Defense, Department of Defense Inspector General, 2009.
[4] G.B. McKenna, Rheology of Roma Plastilina Clay, National Institute of Standards and Technology, Gaithersburg, MD, 1994.
[5] J.E. Seppala, Y. Heo, P.E. Stutzman, J.R. Sieber, C.R. Snyder, K.D. Rice, G.A. Holmes, Characterization of clay composite ballistic witness materials, Journal of Materials Science 50 (2015) 7048-7057.
[6] D. Bhattacharjee, A. Kumar, I. Biswas, V. S., I. E., Thermo-Rheological and Dynamic Analysis of Backing Materials for measurement of Behind Armour Blunt Trauma, Personal Armour Systems Symposium, Amsterdam, 2016.
[7] R. Tao, K.D. Rice, A.M. Forster, Rheology of Ballistic Clay: the Effect of Temperature and Shear History, Society of Plastics Engineers' Annual Technical Conference, Anaheim, CA, 2017.
[8] National Research Council, Phase II Report on Review of the Testing of Body Armor Materials for Use by the U.S. Army, Washington, DC, 2010.
[9] National Research Council, Phase III Report on Review of the Testing of Body Armor Materials for Use by the U.S. Army, The National Academies Press, Washington, DC, 2012.
[10] R. Freeney, R. Mrozek, The Development of a Room Temperature Clay Backing Material for the Ballistic Testing of Body Armor, Personal Armour Systems Symposium, Amsterdam, 2016.
[11] T.D. Edwards, E.D. Bain, S.T. Cole, R.M. Freeney, V.A. Halls, J. Ivancik, J.L. Lenhart, E. Napadensky, J.H. Yu, J.Q. Zheng, R.A. Mrozek, Mechanical properties of silicone based composites as a temperature insensitive ballistic backing material for quantifying back face deformation, Forensic Science International 285 (2018) 1-12.
[12] X. Monnier, J.E. Maigret, D. Lourdin, A. Saiter, Glass transition of anhydrous starch by fast scanning calorimetry, Carbohydrate Polymers 173 (2017) 77-83.
[13] L. Ceseracciu, J.A. Heredia-Guerrero, S. Dante, A. Athanassiou, I.S. Bayer, Robust and Biodegradable Elastomers Based on Corn Starch and Polydimethylsiloxane (PDMS), ACS Applied Materials & Interfaces 7 (2015) 3742-3753.

Tao, Ran; Forster, Aaron; Rice, Kirk; Mrozek, Randy A.; Cole, Shawn; Freeney, Reygan.
"Thermo-Rheological Characterization on Next-Generation Backing Materials for Body Armour Testing."
Paper presented at Personal Armour Systems Symposium 2018, Washington, D.C., United States. October 1, 2018 - October 5, 2018.

First Direct Experimental Studies of $Hf_{0.5}Zr_{0.5}O_2$ Ferroelectric Polarization Switching Down to 100-picosecond in Sub-60mV/dec Germanium Ferroelectric Nanowire FETs

Wonil Chung[1], Mengwei Si[1], Pragya R. Shrestha[2,3], Jason P. Campbell[3], Kin P. Cheung[3] and Peide D. Ye[1]*
[1]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
[2]Theiss Research, La Jolla, CA 92037, USA
[3]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
*Email: yep@purdue.edu

ABSTRACT

In this work, ultrafast pulses with pulse widths ranging from 100 ps to seconds were applied on the gate of Ge ferroelectric (FE) nanowire (NW) pFETs with FE $Hf_{0.5}Zr_{0.5}O_2$ (HZO) gate dielectric exhibiting steep subthreshold slope (SS) below 60 mV/dec bi-directionally. With applied gate bias pulses (V_G = -1 to -10 V), high-mobility Ge drain current was monitored as a test vehicle to capture the polarization switching of HZO. It was found that HZO could switch its polarization directly by a single pulse with the minimum pulse width of 3.6 ns. The polarization switching triggered by pulse train with pulse width as short as 100 ps was demonstrated for the first time.

INTRODUCTION

To further reduce device's SS for lower power operation, negative capacitance FETs (NC-FETs) and use of ferroelectric oxide have been recently studied intensively [1]–[4]. Embedded FE oxides such as HZO within the gate stack of a conventional structure has proven to be successful in SS reduction below the 60 mV/dec limit on various channel materials [5]–[10]. HZO specifically has been reported to be reliable in terms of its switching and polarization capability [8], [11]–[15]. The time response of ferroelectric polarization switching is crucial to evaluate the working speed of HZO based ferroelectric FETs (Fe-FETs) and NC-FETs. However, time response properties of the HZO in FE-FETs were rarely studied. Specifically, at ultrafast sub-10ns regime (towards GHz working frequency), time response of HZO has not been extensively studied yet. In this work, we report the *direct experimental observation* of HZO polarization switching under deep sub-10ns and extend the study using pulse train with 100 ps pulses.

EXPERIMENT

Process flow for the fabrication of Ge FE NW FET with FE HZO gate stack was elaborated in our previous report [6] except the NW release step after fin formation. GeOI wafer was implanted by BF_2^+ ions. Dimensions of fins were controlled by SF_6-based dry etching. HF solution was used to release the Ge NWs from SiO_2. 1 nm Al_2O_3 was deposited by atomic layer deposition (ALD), followed by post-oxidation (O_2, 500 °C) to form ultrathin GeO_x. Subsequent HZO (10 nm) and Al_2O_3 (1 nm) were deposited by ALD and post deposition annealed at 500 °C. Source and Drain areas were etched and deposited with Ni for optimum contacts. 3D structure and SEM images of the final devices are shown in Fig. 1 and 2 respectively.

RESULTS AND DISCUSSION

Firstly, the HZO film was analyzed with XRD as shown in Fig. 3 revealing its non-centrosymmetric, orthorhombic crystal structure which causes the FE property [10]. Ferro-electricity of HZO can be further confirmed with P-V measurement (Fig. 4) showing a clear ferroelectric loop [5]–[7].

Fig. 5 is a typical I_D-V_G curve of the fabricated Ge FE NW pFET. Steep sub-60mV/dec SS for five orders of I_D changes are observed bi-directionally. Fig. 6 (a) and (b) show I_D-V_D curves swept in forward and reverse direction, respectively. Negative differential resistance (NDR) is observed owing to negative drain-induced-barrier-lowering (DIBL) [5]. Fig 7 (a) and (b) depict the instrument set-up for generation and measurement of ultrafast pulses with logarithmically increasing widths from 100 ps to few seconds. Agilent 81110 Pulse Generator (PG) was used with current amplifier and a Lecroy oscilloscope (Fig. 7 (a)) for pulse width > 3.6 ns. For picosecond pulses, as shown in Fig. 7 (b), AVTECH PG was used triggered by the Agilent 81110 PG with a period of 2 µs. Lecroy oscilloscope operating at 80 GS/s sampling rate was used to monitor the sub-ns pulse precisely. V_G and V_D applied to the gate are shown in Fig. 8 (a) and (b). To switch the polarization to off-state, as shown in Fig. 8 (a), initialization pulse V_G = 5 V was applied (>100 ms). Different V_G levels (-1 ~ -10V) were pulsed to the gate with various pulse widths and change in I_D was monitored precisely with oscilloscope in real time through a current amplifier. V_G and V_D were 0 V and -50 mV respectively during the measurements carried out between two adjacent V_G pulses to minimize the effect on the polarization. For lower V_G range (-1 ~ -4 V), Keysight B1530A Waveform Generator/Fast Measurement Unit (WGFMU) was used and the measurement time was fixed at 100 µs. Fig. 9 shows that it takes much shorter time to switch the polarization when the V_G pulse approaches -4 V. As pulse width was decreased to sub-µs regime, transient current fluctuation was observed when the polarization switching occurred as seen in Fig. 10 (a) and (b). Therefore, the current was read after stabilization to ensure it was the current caused by changed polarization state only. At higher V_G (-5 ~ -10 V), sub-10ns switching was also observed directly by experiment and it switched fastest at V_G = - 10 V (Fig. 11). Off-state current was plotted at t = 1 ns to show clear I_D transition from off to on-state. The fastest switching by a single pulse observed in our Ge FeFETs with 10 nm HZO was 3.6 ns (Fig. 12). Considering the RC delay present in the measurement set-up and the large probing pads, the intrinsic time response of polarization is expected to be faster than 3.6 ns. To further investigate the effect of sub-ns pulses, Fig. 7 (b) set-up was configured and 100 ps pulses with period of 2 µs (Fig. 13 (a), V_G = -6 V) were generated in the form of pulse train (duty cycle = 0.005 %). Although a single 100 ps pulse (Fig. 14) did not trigger noticeable polarization switching, as the pulses accumulated, polarization switching could be detected (Fig. 15). Further studies on the effect of duty cycle with picosecond pulse widths could be valuable for deeper understanding of dynamics of polarization switching in HZO devices for various applications including FeRAM.

CONCLUSION

High mobility Ge FE NW pFETs were applied as test vehicles for the first time to study the time response of FE HZO gate stack down to 100 ps. It was found that a single deep sub-10ns pulse or even a pulse train with 100 ps pulses were enough to initiate polarization switching in HZO. This work opens the route to studying the dynamics of FE HZO through direct experiments down to 100 ps and even beyond. The work is supported by SRC and Lam Research. U.S. Government is not endorsing any of the equipment mentioned in the paper.

References: [1] S. Salahuddin et.al., *Nano Lett.,* 2008. [2] Z. Krivokapic et.al., *IEDM*, 2017. [3] H. Ota et. al., *IEDM*, 2016. [4] P. Sharma et.al., *VLSI*, 2017. [5] M. Si et. al., *Nat. Nanotechnol.*, vol. 13, p. 24–28, 2018. [6] W. Chung et. al., *IEDM*, 2017. [7] M. Si et. al., *IEDM*, 2017. [8] M. H. Lee et. al., *IEDM*, 2016. [9] C.-J. Su et. al., *VLSI*, 2017. [10] J. Muller et. al., *Nano Lett.*, 2012. [11] K.-Y. Chen et. al., *VLSI*, 2017. [12] H. J. Kim et. al., *Nanoscale*, vol. 8, p. 1383, 2016. [13] J. Muller et. al., *EDL*, vol. 33, no. 2, p. 185-187, 2012. [14] Y. Chiu et. al., *IRPS*, 2015. [15] S. Oh et. al., *EDL*, vol. 38, no. 6, p. 732-735, 2017.

Chung, Wonil; Si, Mengwei; Shrestha, Pragya; Campbell, Jason; Cheung, Kin; Ye, Peide.
"First Direct Experimental Studies of Hf0.5Zr0.5O2 Ferroelectric Polarization Switching Down to 100-picosecond in Sub-60mV/dec Germanium Ferroelectric Nanowire FETs."
Paper presented at 2018 Symposia on VLSI Technology and Circuits, Honolulu, HI, United States. June 18, 2018 - June 22, 2018.

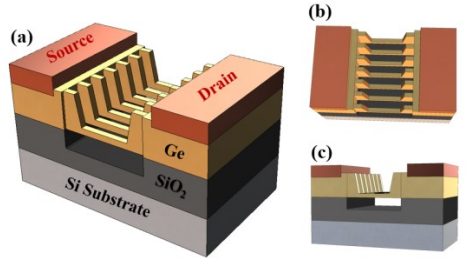


Fig. 1. 3D structure of a Ge NW FET viewed from **(a)** top-right, **(b)** top and **(c)** front. Under the nanowires, SiO_2 was etched and kept hollow.

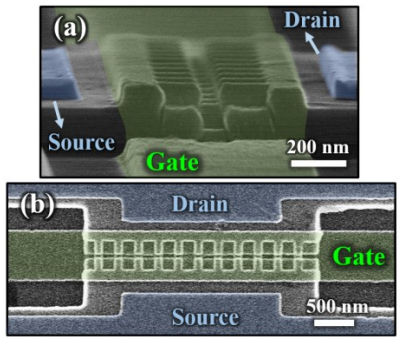


Fig. 2. SEM images of fabricated device viewed from **(a)** side and **(b)** top. 11 parallel NWs make one device.

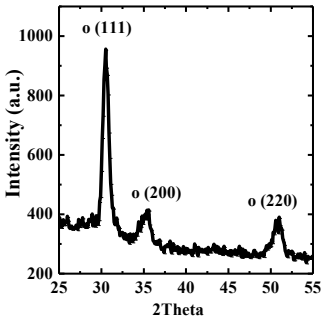


Fig. 3. XRD peaks of the $Hf_{0.5}Zr_{0.5}O_2$ film used in the device exhibits orthorhombic structure.

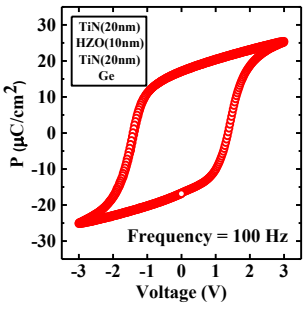


Fig. 4. Polarization-Voltage (P-V) graph of 10 nm HZO film measured at frequency of 100 Hz.

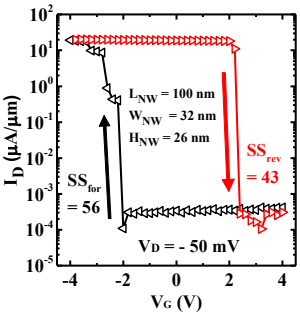


Fig. 5. I_D-V_G exhibits bi-directional SS < 60 mV/dec for 4~5 orders of I_D changes.

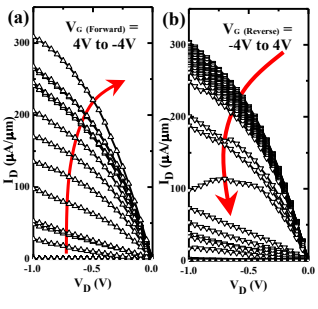


Fig. 6. I_D-V_G of **(a)** forward and **(b)** reverse sweeps. Negative Differential Resistance (NDR) can be observed in reverse sweep direction.

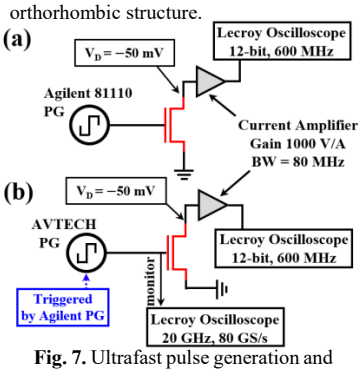


Fig. 7. Ultrafast pulse generation and measurement set-up for **(a)** pulse widths > 3.6 ns and **(b)** picosecond pulses.

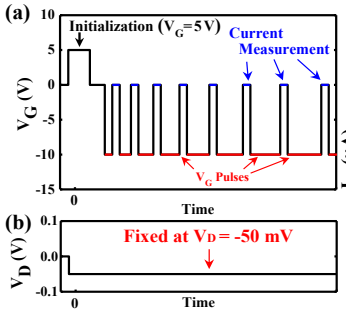


Fig. 8. Pulse time line of **(a)** V_G and **(b)** V_D. Currents were measured in between pulses with V_G = 0 V.

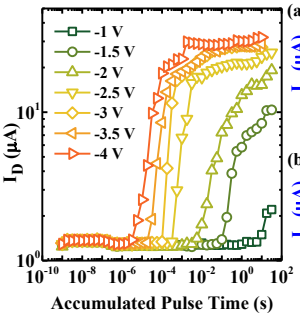


Fig. 9. Polarization current (I_D) over accumulated pulse time with different V_G pulses. Keysight B1530A was used for these pulses.

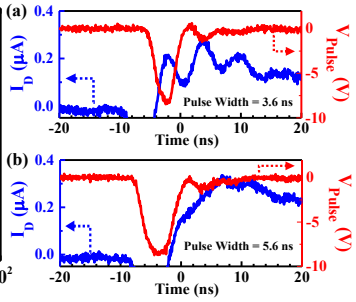


Fig. 10. Polarization switching can be monitored by I_D. Time responses with pulse widths of **(a)** 3.6 ns and **(b)** 5.6 ns are shown. I_D was taken after the fluctuation was stabilized.

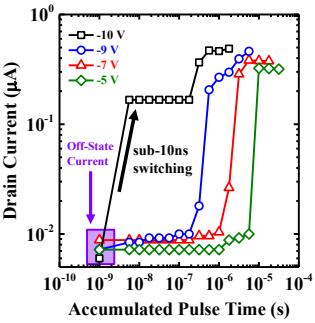


Fig. 11. V_G was pushed down to -10 V to probe the polarization switching limit which showed minimized time under 10 ns. Off-state current was plotted at 1 ns for reference.

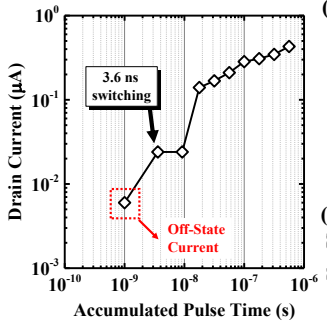


Fig. 12. The fastest switching was observed at accumulated V_G pulse time of 3.6 ns. Off-State current was plotted at 1 ns for reference.

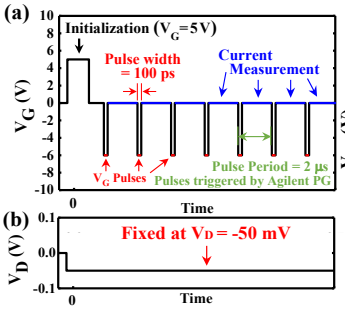


Fig. 13. (a) V_G and **(b)** V_D settings for sub-ns pulse measurements with Fig. 7 (b) configuration. I_D was kept constant at -50 mV and 100 ps pulses were triggered every 2 μs (duty cycle = 0.005 %).

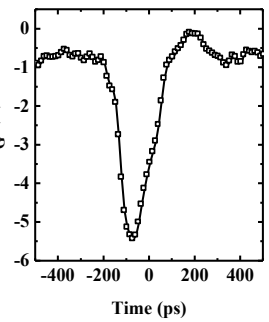


Fig. 14. 100 ps pulse generated with rise time of 60 ps was measured at the rate of 80 GS/s. Maximum voltage was -6 V. To avoid loss of voltage due to cables, the PG was placed very close to the device.

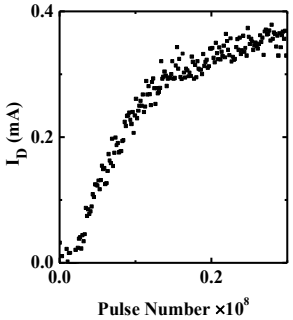


Fig. 15. Polarization switching was monitored by applying the 100 ps pulse shown in Fig. 14 in the form of a continuous pulse train.

Chung, Wonil; Si, Mengwei; Shrestha, Pragya; Campbell, Jason; Cheung, Kin; Ye, Peide.
"First Direct Experimental Studies of Hf0.5Zr0.5O2 Ferroelectric Polarization Switching Down to 100-picosecond in Sub-60mV/dec Germanium Ferroelectric Nanowire FETs."
Paper presented at 2018 Symposia on VLSI Technology and Circuits, Honolulu, HI, United States. June 18, 2018 - June 22, 2018.

2019 IEEE Conference on Communications and Network Security (CNS)

BotSifter: An SDN-based Online Bot Detection Framework in Data Centers

Zili Zha
George Mason University
zzha@gmu.edu

An Wang
Case Western Reserve University
axw474@case.edu

Yang Guo, Doug Montgomery
NIST
{yang.guo, dougm}@nist.gov

Songqing Chen
George Mason University
sqchen@cs.gmu.edu

***Abstract*—Botnets continue to be one of the most severe security threats plaguing the Internet. Recent years have witnessed the emergence of cloud-hosted botnets along with the increasing popularity of cloud platforms, which attracted not only various applications/services, but also botnets. However, even the latest botnet detection mechanisms (e.g., machine learning based) fail to meet the requirement of accurate and expeditious detection in data centers, because they often demand intensive resources to support traffic monitoring and collection, which is hardly practical considering the traffic volume in data centers. Furthermore, they provide little understanding on different phases of the bot activities, which is essential for identifying the malicious intent of bots in their early stages.**

In this paper, we propose BotSifter, an SDN based scalable, accurate and runtime bot detection framework for data centers. To achieve detection scalability, BotSifter utilizes centralized learning with distributed detection by distributing detection tasks across the network edges in SDN. Furthermore, it employs a variety of novel mechanisms for parallel detection of C&C channels and botnet activities, which greatly enhance the detection robustness. Evaluations demonstrate that BotSifter can achieve highly accurate detection for a large variety of botnet variants with diverse C&C protocols.

I. Introduction

Driven by the great success of cloud computing, the last decade have seen the continuous migration of various services and applications to different cloud platforms. With the flexible pay-per-use model, more and more end users also rely more and more on the cloud platforms for their personal storage and computing need [1]. Under this trend, bots and botnets-as-a-service [2] are no exception: bots and botnets have been one of the most severe Internet threats underpinned by the economic motive, and recent years have witnessed the emergence of cloud-hosted botnets [3] largely due to its cost effectiveness and the long-term availability of the machines in the data centers. Compared to traditional bot machines which get switched on/off frequently by the end-users, data center hosted bot machines usually stay online for longer periods of time [3] and generate more attack traffic (and profit). It was previously reported [4] that cybercriminals have managed to install DDoS botnets in AWS by exploiting a vulnerability in Elasticsearch [5], an open source search engine that is often deployed in cloud environments such as Amazon EC2, Microsoft Azure, Google's Compute Engine. Likewise, botnet command and control software has previously been found to be hosted in Dropbox [6]. Apart from these particular cases, they show that cybercriminals are now breaching commodity data centers, seeking the values of cloud infrastructures to host the botnets.

Cloud-hosted bots (and thus truly botnets-as-a-service) are more harmful than their traditional counterpart, yet accurate and expeditious botnet detection schemes on cloud platforms are not on the horizon yet because of the following challenges. First, since the cloud is a multi-tenant environment, the detection of bots should be fast (e.g., at runtime to prevent further damage) and as non-intrusive as possible so that the detection would have no or trivial impact on the normal data center applications. Second, given the large traffic volume in a data center, such a detection scheme must be scalable, capable of handling tens or hundred gigbit line rate. Third, the detection must be very accurate, since compared to the end user environment, the cost of misclassifying a bot process or connection (and subsequently closing/blocking the connection) in a data center could be much larger or even disastrous.

Some of the early-day bot detection schemes [7] [8] rely on deep packet inspection, which suffer from high resource demands and ineffectiveness against encrypted botnet traffic. Some others [9], [10], [11], [12], [13] focused on host traffic and bot behavior analysis. These schemes became less and less effective since contemporary botnets are constantly evolving to circumvent the advanced detection mechanisms [14], [15], [16], [17], [18]. For example, most botnets abandoned the traditional IRC based Command and Control (C&C) channels and embrace HTTP or P2P for communications.

To deal with such challenges, lately more schemes have been built by taking advantage of the machine learning (ML) techniques for detecting bot activity and/or C&C channels [9], [10], [18], [12], [13], [19]. Such schemes often demand intensive resources to capture the incoming and outgoing traffic information, e.g., a centralized traffic monitoring facility at the network gateway or the firewall, and then run the detection after the traffic features are extracted. These schemes may work for a small network with medium or low traffic volume, but they are hardly effective in detecting the cloud-based bots because (1) they demand a lot of extra resources for effective traffic monitoring, which is hardly scalable considering the huge traffic volume in data centers, (2) their accuracy varies with the used ML model and the chosen features, and they are often incapable of identifying unseen bots without understanding bot invariant characteristics (e.g., C&C channels), and (3) they provide little understanding on different phases of bot

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

activities, which is essential for identifying the malicious intent of bots in their early stages. This is however very desirable for data centers in order to prevent further damage.

More importantly, existing ML-based approaches focus on the detection of botnet flows [12], [13]. However, in the detection of cloud-hosted bots, accurate identification of the bots is more imperative compared to the detection of individual malicious flows. Only after the bots are identified, the infected VMs could be shut down to prevent future attacks. To our best knowledge, no prior works focused on the detection of VM-based bots in the cloud.

To this end, we propose to build BotSifter towards a scalable and accurate runtime bot detection framework in data centers. To be scalable, BotSifter integrates centralized learning (thus to have a global view of the traffic and centralized intelligence) with distributed edge-assisted detection by leveraging software switches in data centers. Due to their widespread deployment in data centers, software switches (e.g., Open vSwitch[1]) are being increasingly employed as monitoring devices as they typically reside in commodity servers with abundant hardware resources. Since they are usually deployed at the edge of the monitored network and located within close proximity to the end hosts, only relatively a small amount of traffic flows traverse the switch, rendering it a more scalable solution for anomaly detection in data centers. Implementing detection at the edge also enables our system to observe both directions of the traffic, which is an essential prerequisite for connection based anomaly detection. Moreover, our edge-based detection framework has the inherent capability of observing the internal attack traffic and C&C traffic within the data center network.

To achieve high accuracy, BotSifter not only conducts neural network (NN) based bot activity detection, but also conducts the detection of C&C channels in parallel. These detections are further enhanced with local and network wide correlations to minimize false alarms. Not only detecting the existence of C&C channels, BotSifter is also able to differentiate different communication protocols (i.e., IRC, HTTP, P2P) utilized by the bots so that custom mitigation mechanisms could be quickly deployed to defend against specific types of bots. Since the majority of the detection operations in BotSifter are conducted within software switches that are instrumented to collect connection features and states on the fly, BotSifter is highly efficient in identifying bots without interrupting other network functions, thus making it a practical solution for the cloud and data center systems. A prototype of BotSifter is implemented, and evaluations based on real-world traces show BotSifter is highly accurate and efficient in detecting bots utilizing known protocols, but also bots with customized protocols.

The highlights of BotSifter lies in

- The design naturally utilizes the SDN's centralized structure to have a global visibility while distributing most of the detecting load to the network edge to be scalable.
- It builds the monitoring capability in the OVS via an off-path traffic collection design to minimize the intrusion of traffic monitoring to normal applications.
- It utilizes neural network to detect bot activities, and employs newly designed protocol specific mechanisms (e.g., self-correlation for P2P) to detect and differentiate different C&C protocols. The local and network wide correlations further enhance the accuracy and robustness of the detection.

The remainder of the paper is as follows. Section II describes the BotSifter design and section III sketches its implementation. The evaluation is presented in section IV. We discuss the related work in section V and make concluding remarks in section VI.

II. BotSifter Design

The design of a cloud botnet detection framework that is capable of monitoring thousands of servers, tens of thousands of Virtual Machines (VMs) running on these servers, and terabit per second communication among these entities and between the data center and the outside Internet is extremely challenging. In this paper, we explore a ML based distributed online cloud bot detection framework, called BotSifter, that monitors and detects the bots within a data center. BotSifter takes advantage of Software Defined Networking (SDN) architecture widely adopted by the data center, and integrates the centralized ML training with distributed monitoring and detection. The ML based bot detector is trained at the central controller, and runs distributedly at the cloud servers with the ML configurations acquired from the central controller. SDN software switches, e.g., OVS, are instrumented with traffic monitoring capability. Both bot activity detector and bot C&C communication detector are implemented at servers using locally collected traffic stats. If necessary, the central detector is invoked to detect data-center wide botnets.

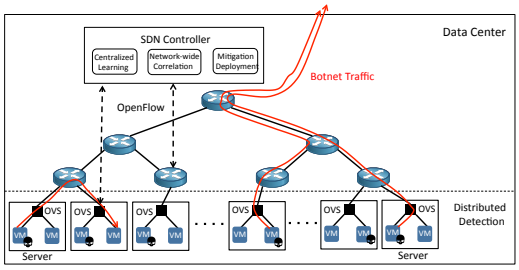


Fig. 1: Overall architecture of BotSifter

The key feature of our design is to leverage the data-center servers and the software switches running on these servers. The contemporary servers have abundant computational and memory resources, and the software switches have the full access to the traffic originated from and destined to the end hosts residing on the servers. Our design philosophy is to

[1]Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

place as many monitoring and detection functionality at the edge servers as possible, while judiciously utilize the central controller for data-center wide tasks when necessary.

The BotSifter architecture is depicted in Fig. 1. The central controller is responsible for ML training, network wide bot detection, and bot mitigation. At individual servers, traffic features are extracted and recorded by the instrumented software switches, which are then used by C&C channel detection modules and ML based bot activity detection module. Below we describe local traffic monitoring, C&C bot detection, ML based bot detection, and network wide bot detection and mitigation, respectively. The major components to accomplish these tasks are sketched in Figure 2.

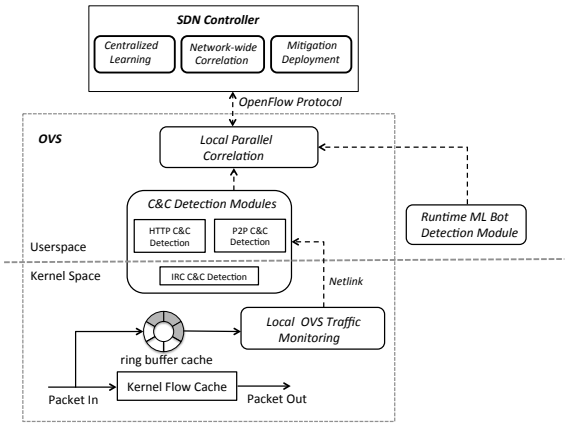


Fig. 2: BotSifter design: major components

A. Local OVS Traffic Monitoring

To lessen the impact on software switch's forwarding speed while efficiently capturing the traffic stats required by ML bot detection module and C&C detection modules is the main design challenge for local OVS based traffic monitoring. OVS maintains a user-space forwarding pipeline and a kernel space forwarding cache. The majority of the incoming packets are forwarded by the kernel module, with few packets that do not have the matches in the kernel being redirected to the user-space. To decouple the monitoring from the forwarding, a ring buffer cache, as shown in the bottom of Fig. 2, is introduced. Such a design significantly reduces the monitoring interference to the forwarding.

ML detection module and different C&C detection modules require different traffic stats. For instance, ML module uses a traffic feature vector, while P2P C&C detection keeps track of DNS transactions for each source/destination pair. Multiple hash tables are employed for efficient lookup and update.

B. Parallel C&C and Bot Activity Detection

At edge server, BotSifter implements one ML based bot detection module and three C&C detection modules: HTTP C&C detection module, P2P C&C detection module, and IRC C&C detection module, respectively. We have the design choice of implementing them inside the software switch OVS, or on the server but outside the OVS. In addition, if a module is placed inside the OVS, we need to decide whether the module resides in the kernel or at user-space. Since P2P and HTTP C&C modules use the connection stats rather than individual packet info, placing them in the user-space at OVS is the right choice. In contrast, IRC based C&C module conducts keyword search over a packet, thus is impleneted in OVS kernel. ML based bot detector uses TensorFlow ML libraries, thus is implemented in the user-space outside OVS. We also place the local parallel correlation module in the OVS, as shown in Fig. 2.

1) C&C Detection: **P2P C&C detection:** Distinguishing P2P traffic used by bots from normal P2P traffic is not trivial. The study in [11] made the discovery that the sets of peers contacted by two different bots within the same botnet typically have a much larger overlap compared to the peer sets contacted by two legitimate P2P clients within the same network. In practice, there is the possibility that only one bot resides in the data center.

We develop a client self-correlation approach to detect P2P bots. Specifically, the peer set of a P2P bot is more likely to remain stable over the time, while that of a normal P2P client is more likely to change due to the changing user behavior (e.g., downloading of disparate resources over time). We conduct extensive empirical experiments, and the results show that P2P bots do exhibit strong self-overlap patterns while normal P2P hosts do not. The new approach is more flexible and robust than the approach in [11].

HTTP C&C detection: Bots tend to communicate with the server periodically over the HTTP channel [20]. HTTP C&C detection module makes the detection using such periodic traffic pattern. The timing information of HTTP connections between a pair of end hosts is collected by the monitoring module. The number of HTTP connections within each time slot is counted. The time series of connection counts is fed into a Discrete Fourier Transform (DFT). If periodic pattern is discovered, the host becomes a bot candidate.

IRC based C&C detection. Although IRC based botnets are diminishing nowadays, we include the IRC C&C detection for the sake of completeness. To minimize system overhead, a combination of keyword matching and port numbers is leveraged to identify IRC connections. Furthermore, to determine whether they are IRC C&C channels, our detection relies on inspection of packet payloads by searching for attack relevant commands in IRC response messages. We implement the detection module at the kernel space of OVS.

Note that our system design is extensible and new C&C detection schemes can be easily integrated. In addition, the design enables network wide correlation for agile C&C detection. For example, a P2P C&C host may not exhibit sufficient self-overlap. In such a case, flow stats of all P2P hosts inside a network can be exported to the central controller, where the clustering based detection can be performed.

2) Runtime ML based bot detection: In parallel to the C&C channel detection, BotSifter also conducts bot activity detection using a deep learning Neural Network (NN). For

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

each connection, NN is able to detect if it is potentially a bot activity connection.

In order to detect if a host is compromised and becomes a bot, BotSifter keeps track of the percentage of connections initiated by this host that are identified by the NN as the bot activity connections. If the percentage surpasses a preset threshold, the host is identified as a bot.

3) Local parallel correlation: While C&C based detection detects bots via monitoring C&C communication, ML based detection detects the bots via monitoring bot activity traffic. BotSifter introduces a local detection correlation module that combines the results from these two types of detection modules. The consistent detection by both C&C modules and ML based module is a strong indication that a host has been compromised and will be placed on a blacklist. On the other hand, a single positive identification may not be conclusive. For example, the P2P C&C detection module detects that a host may potentially be a bot. However, the bot may still be dormant and have not launched any attacks yet. As another example, if a host is identified by the online ML module as an attacker, it is not clear if the identified host is a bot with a customized C&C protocol that is not recognized by our C&C detection modules, or an ordinary attacker without any C&C channels. In any case, such bots are placed on a so-called local grey list and will be under persistent monitoring/scrutiny. The blacklists and grey lists are sent to the central controller for network-wide correlation and identification.

C. Network-wide Correlation and Mitigation

While local detection is beneficial for the system scalability, the lack of global view can deter the bot detection. Hence BotSifter introduces a centralized correlation module that examines the blacklists and greylists received from the edge servers, and performs network-wide correlation to detect bots. In addition, a mitigation module is implemented to mitigate bots' damages. For each bot on the blacklist, the controller installs flow rules into the OVS to block C&C traffic and/or attacking traffic.

For hosts on a local C&C greylist, the controller performs network-wide grouping analysis to determine if it is a bot. Each group consists of the hosts on blacklists or C&C greylists communicating with the same destination host. If any host in the same group has already been positively identified as a bot, the greylist hosts in the same group are marked as a bot and will be placed on the blacklist. Otherwise, the controller places the host on the global C&C greylist, and continues to monitor the host until a timeout occurs. For hosts on the local ML greylist, the controller looks up the global C&C greylist to check if there is a match. If so, this host is included in the global blacklist. Otherwise, it is placed onto the global ML greylist and continues to be monitored for future signs of C&C communication.

III. BotSifter Implementation

We implement a BotSifter prototype following the design as laid out in the Section II. Fig. 3 highlights the major parts implemented for BotSifter. More implementation details are described as below.

A. Local OVS Traffic Collection Implementation

The local traffic monitoring function is implemented in a kernel thread called *kernel_collector*. We also modify the kernel forwarding thread so that the packet headers are pushed into the ring buffer cache upon arrival. The *kernel_collector* thread takes the packet headers off the ring buffer, and process them to generate the required stats that are stored in hash tables. The hash tables with connection stats are periodically pulled by the user-level stats collection thread *stats_collector*, as shown in Fig. 3.

A more involved task is to detect P2P connections required by the P2P C&C detection module. To identify a P2P connection, the kernel thread intercepts the DNS requests, parses them, and records the hosts who have conducted look-up for a specific IP address. For each new connection between any two hosts *src_ip* and *dst_ip*, we check whether the *src_ip* host has conducted a DNS lookup for the destination *dst_ip*. If yes, it is a normal connection. If not, this new connection is marked as a P2P connection, which will be further examined by the P2P C&C detection module.

B. Parallel C&C and Bot Activity Detection Module Implementation

Three threads, *irc_c&c_detector*, *p2p_c&c_detector* and *http_c&c_detector* are implemented to realize the IRC C&C detection, P2P C&C detection, and HTTP C&C detection, respectively. Thread *irc_c&c_detector* runs in the kernel, as discussed in the Section II, while *p2p_c&c_detector* and *http_c&c_detector* run in the user space, as in Fig. 3.

To identify IRC connections, *irc_c&c_detector* performs lightweight payload inspection against connections with standard IRC ports (6660-6669). The first few packets of an IRC connection typically contain certain keywords, such as "JOIN","USER", "NICK" and "PRIVMSG". To further identify IRC C&C channels, *irc_c&c_detector* examines IRC messages by searching for attack relevant commands (e.g., "scan", "flood"). If such commands are recognized, the *irc_flag* for the connections will be set and exported to userspace.

As mentioned in Section II, the differentiation of P2P C&C traffic and normal P2P traffic relies on self-correlation behavior of peer sets. For the self-correlation, we conduct overlap analysis over multiple time windows of each P2P connection. Time window is a fixed length period of time during the connection. For each time window, we calculate the per-host peer sets in the current and N subsequent time windows. Then, we calculate the number of re-appearing peers in both the current time window and the $i^{th}(i \leq N)$ time window. We use the average of the N overlap values to represent how the peer sets evolve over time. For runtime detection, we calculate the moving average overlaps to differentiate the P2P bots from normal hosts. The used parameters are discussed in Section IV.

Thread *online_ml_detector* implements a NN based bot detection module. The NN model is implemented using TensorFlow 1.8.0 [21], an open source NN library. Since the thread

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

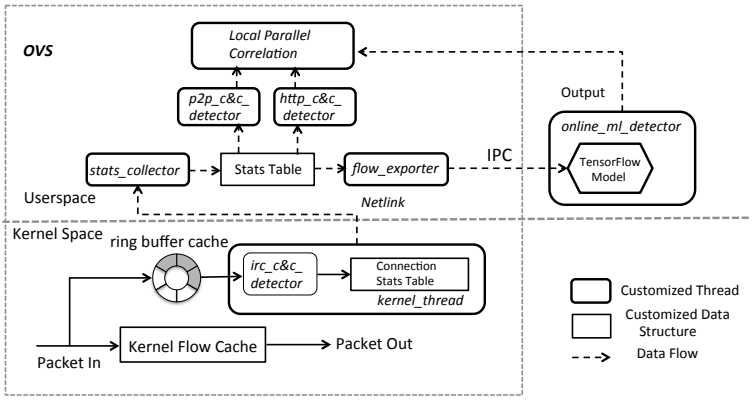


Fig. 3: BotSifter system implementation

runs outside of the OVS, a user-space thread inside OVS, called *flow_exporter*, is created to allow *online_ml_detector* to retrieve the connection [2] feature vectors collected by OVS using Linux IPC calls, as depicted in Fig. 3.

Table I summarizes the features used in our NN model. Notably, different from previous flow based ML models, *src/dst* IP addresses are excluded from our feature set. We believe they are unique to particular data centers and thus should be excluded. More importantly, the exclusion can avoid over-fitting the NN model. In general, choosing the right feature set, or feature engineering, is a challenging research task. We experimented with different features and choose the most discriminative ones based on the experiment results. For example, the forward/backward average packet size of normal flows tends to be larger than botnet traffic since they contain realistic payloads. Meanwhile, we introduce a novel feature, i.e., *conn_stat* to capture whether the connection has been successfully established. Since bots usually maintain a large number of half open connections, this feature is effective in identifying malicious botnet traffic such as direct DoS flooding attack and the amplification attack traffic.

The *online_ml_detector* thread fetches NN configuration parameters from the controller, and initiates the NN model. The *flow-exporter* thread in OVS continuously exports connection features to the NN model, as shown in Fig. 3. To facilitate communication between *flow_exporter* and *online_ml_detector*, a shared memory pool is created. Linux semaphores are used to synchronize access to the shared memory.

C. Network-wide Correlation and Mitigation Implementation

Bot detection applications are programmed at the central controller to perform network-wide correlation analysis. Once a bot is detected, the controller installs customized flow rules into the corresponding OVS switch to prevent future attacks. To facilitate the communication between the threads at servers and controller, traditional OpenFlow protocol is extended to facilitate the collection of detection results from edge servers.

[2]A UDP "virtual" connection is considered successful if at least one packet is received from the destination host.

TABLE I: Description of connection features.

Feature	Description
Source Port	Source port.
Destination Port	Destination port.
Protocol	IP protocol.
Forward packet count	Total number of packets in the forward direction.
Backward packet count	Total number of packets in the backward direction.
Forward byte count	Total number of bytes in the forward direction.
Backward byte count	Total number of bytes in the backward direction.
Duration	Connection duration.
Forward packet rate	Packet rate in the forward direction.
Backward packet rate	Packet rate in the backward direction.
Connection State	whether the connection is successfully established.
Forward inter-arrival time	Packet inter-arrival time in the forward direction
Backward inter-arrival time	Packet inter-arrival time in the backward direction.
Forward average packet size	Average packet size in the forward direction.
Backward average packet size	Average packet size in the backward direction.
Forward byte rate	Byte rate in the forward direction.
Backward byte rate	Byte rate in the backward direction.

IV. Evaluation

Our testbed consists of three Lenovo ThinkServer machines running Ubuntu 14.04. Each machine is equipped with Intel Xeon 4-Core 3.20GHz CPU and 4GB RAM. One machine is installed with Open vSwitch 2.3.90. Another machine runs the Ryu SDN controller. A third machine serves as both packet generator and data sink, which is connected to the OVS machine via two 10Gbps Ethernet cables. Packet traces are replayed using TCPReplay at the original speed.

A. Datasets

For our experiments, we are able to obtain several third-party traces, including a variety of botnet traces composed of different botnet types and normal network traces. CTU-13 dataset [22] contains traffic of botnet samples captured in the CTU University, Czech Republic, in 2011. In each botnet sample, bots use specific protocols for C&C communication and perform diverse malicious tasks, including SPAM, port scanning, DDoS, click fraud, etc. Due to the diversity of botnet samples, this dataset is appropriate for evaluating the performance of our framework. For the normal traces, we use the UNB ISCX IDS 2012 dataset [23] containing normal traces with full packet payloads. This dataset captures typical user daily activities (e.g., HTTP, DNS, SSH, FTP) at UNB university and was made public for scientific research.

Our runtime bot activity detection relies on a pre-trained model. Table II summarizes the composition of the botnets in

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

Datasets	*Underlying C&C Protocol*			
	HTTP	***P2P***	***IRC***	***Custom***
Training Dataset	Neris, Virut	Storm, Waledac	Rbot	-
Testing Dataset	Neris, Virut, Sogou	Storm, Waledac, NSIS	Rbot	Menti

TABLE II: Description of botnet datasets for training and testing.

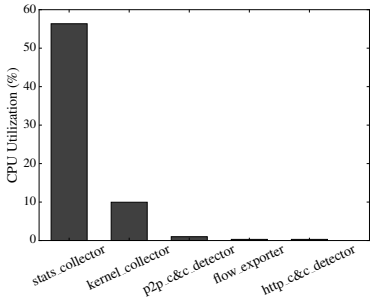


Fig. 4: CPU utilization of detection related threads.

the training and testing datasets. For the training dataset, we merged several traces, including P2P botnets, IRC Botnets and HTTP botnets. For runtime detection, we include not only the same types of botnets as the training dataset, but also novel botnet variants, such as NSIS (P2P C&C), Menti (Proprietary C&C) and Sogou (HTTP C&C). Among them, Menti uses a custom protocol for C&C communication. By evaluating how BotSifter performs when facing novel botnet variants, this experimental strategy aims to demonstrate the effectiveness of our design in real-world bot detection.

B. Impact on Normal Applications

Since the botnet detection accuracy of our system strongly relies on the accurate timing information of the packet sequences, the botnet traces are replayed at the original speed. To evaluate the overall system performance, we perform a stress test using a CAIDA trace and replay the trace towards our monitoring system at the highest achievable rate.

Considering that no modification is made for native OVS threads (e.g., handlers and revalidators), their CPU usage is not shown here. We only report the CPU utilizations of the customized threads in Fig. 4, which shows the peak CPU usage of each detection-related thread under various detection tasks.

We can see that the *stats_collector* in the userspace incurs the highest CPU utilization while the threads regarding C&C detection and flow exportation to the online ML detector introduce negligible overhead. Since *stats_collector* mainly manages the collection of connection stats from the kernel space through Netlink and the maintenance of the connection hash table in the userspace, we infer that its CPU utilization is mainly attributed to the Netlink communications. Further optimization is possible by utilizing shared memory between the user and kernel space instead of relying on Netlink sockets.

In realistic scenarios, these overheads are acceptable as OVS typically resides in commodity servers with abundant CPU resources. The detector threads could be pinned to different CPUs to avoid interfering with CPUs dedicated to the forwarding functions in OVS.

To estimate the network overhead, we use DPDK based packet generator MoonGen to generate high speed traffic and measure the maximally achievable throughput such that no packet loss occurs on the forwarding path. The experiment is repeated 10 times. Compared to the throughput of the native OVS (1.44Mpps), BotSifter could achieve 1.15Mpps throughput. This overhead is acceptable and further analysis shows this overhead is mainly incurred by the *memcpy* operations on the ring buffer cache.

C. Detection Performance

1) Evaluation metrics: To evaluate the performance of Botsifter, we use three metrics. *C&C Accuracy* is calculated as the ratio between the number of hosts which have triggered alarms of C&C detectors and the total number of hosts. *ML Accuracy* represents the detection rate of the bots based on suspicious ratios predicted by the runtime NN model. Since our final detection result relies on a parallel correlation between C&C detection and runtime ML detection, we define a novel metric *Local Parallel Accuracy* to represent the overall accuracy.

As discussed in Section II, the hosts triggering alarms from both *http_c&c_detector* and *irc_c&c_detector* will be put onto a greylist instead of blacklist since they need further network-wide correlation. In contrast, alarms from *p2p_c&c_detector* indicate the associated hosts must be bots and thus they are included in a blacklist. Therefore, we claim that if the host appears on either the C&C blacklist or the ML greylist (e.g., is an attacker, but needs further monitoring to find out C&C channels), it is regarded as a true positive, which also implies that it has been successfully detected by BotSifter. This design principle further demonstrates the robustness of BotSifter compared to methods based on single stage detection.

2) Detection Accuracy and Latency: The measurement results are shown in the C&C accuracy column in Table III. Note that, in ML Accuracy, the percentage values in the brackets represent the suspicious ratios for the bots in the current trace. The parallel detection scheme in BotSifter successfully detects all bots with zero false positives, although some bots trigger alerts from only one stage (either C&C or NN model), such as NSIS and Menti. In the normal trace, there are no false positives due to the following two reasons. First, the hosts are included in HTTP greylist instead of the blacklist, due to the periodicity of software updates. Besides, the suspicious ratios of all hosts are far below the threshold. In a nutshell, BotSifter can detect all bots which may get missed by any single stage of detection. Since local detection accomplishes the tasks, no computation is needed from the central controller, which demonstrates that BotSifter is a distributed scalable solution. In the following, we analyze the results in more detail.

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

TABLE III: Detection results for 8 botnet variants and normal trace.

Botnet Variant	C&C Protocol	#Bots or #Hosts	Duration	P2P Overlap	HTTP Periodicity	IRC C&C	C&C Accuracy	ML Detection Accuracy	Local Parallel Accuracy
Neris	*HTTP*	1	4.8*h*	-	*Strong*	-	1/1	1/1(99.2%)	1/1
Virut	*HTTP*	1	16.36*h*	-	*Strong*	-	1/1	1/1(99.7%)	1/1
Sogou	*HTTP*	1	0.38*h*	-	*Weak*	-	1/1	1/1(95.5%)	1/1
Storm	*P2P*	13	3.1*h*	*Yes*	-	-	13/13	13/13(90.1%)	13/13
Waledac	*P2P*	3	3.45*h*	*Yes*	-	-	3/3	0/3(50.5%)	3/3
NSIS	*P2P*	3	1.21*h*	-	-	-	0/3	1/1(93.7%)	3/3
Rbot	*IRC*	1	5*h*	-	-	Yes	1/1	1/1(99%)	1/1
Menti	*Custom*	1	2.18*h*	-	-	-	0/1	1/1(99.4%)	1/1
Normal	*n/a*	36	10*h*	-	*Yes(4/36)*	-	4/36 on greylist	FPR:0/36(See Fig. 6)	FPR: 0/36

HTTP C&C detection. Based on our empirical experiments, we find that the C&C connections of certain HTTP botnets may not exhibit periodicity patterns as strong as other botnets. This may be due to unknown factors such as network delay and congestion which introduce noises to the connection timing. We use *strong* and *weak* to represent whether strong periodicity is observed in each botnet trace. Among the used HTTP botnet samples, Neris and Virut exhibit strong periodicity since the bots connect to the HTTP C&C server at regular intervals with negligible timing variations. Since periodicity patterns are observed for each 3-tuple *(src_ip, dst_ip, dst_port)*, the suspicious client and server associated with the 3-tuple could be accurately pinpointed and placed onto a greylist for further examination. Aside from this, we also observed that certain botnets contain more than one HTTP C&C channels, some of which exhibit no periodical pattern. These C&C communications are not typical and cannot represent the botnet C&C behaviors. For those C&C channels exhibiting periodical connection patterns, our online *http_c&c_detector* can accurately identify the C&C channels.

Apart from accuracy, the detection latency is dominated by the prevalence of bots' activities. As previously discussed, HTTP C&C detection relies on disclosing the periodicity inherent in bot C&C communication. Therefore, the latency for identifying the C&C channels is closely correlated with the inter-connection duration. In our experiment, BotSifter can detect periodic C&C communication after 4∼5 connections on average. The actual latency depends on how frequent bots connect to C&C servers. For example, in one of our Neris botnets, the bot periodically communicates with the C&C server on a minute basis, in which BotSifter can detect this channel after around 4∼5 minutes since the first connection.

During the test for the normal trace, the *http_c&c_detector* triggers an alert unexpectedly, which implies that periodic http connections are observed between two normal hosts. Further examination reveals that those hosts are running legitimate applications with periodic HTTP connections.

P2P C&C detection. The detection results for the P2P botnets are also quite promising. Storm and Waledac both show strong self-overlap in their C&C communication. One exceptional case is NSIS, for which no P2P overlap is observed. This is reasonable as the trace is reported to be incomplete with only one C&C channel that is insufficient for analysis. Below we demonstrate that our scheme allows to accurately distinguish P2P C&C from normal P2P traffic.

TABLE IV: Description of P2P traces.

Trace	*Normal*			*Botnet*	
	Emule	*FrostWire*	*uTorrent*	*Storm*	*Waledac*
Number of Packets	3.6M	1.35M	6.5M	6M	2.5M
Duration	3.25*h*	10*h*	8*h*	3.16*h*	8.23*h*

As previously discussed, the key characteristic differentiating P2P C&C from normal P2P traffic is that the peer sets of a P2P bot have large overlaps across consecutive time windows; while the overlaps for a normal P2P host are noticeably smaller. To verify this, we use several third-party traces [24] of normal P2P applications and P2P botnets, as summarized in Table IV. The Storm and Waledec contain 13 and 3 P2P bots respectively. Our experiments show that the *p2p_c&c_detector* could successfully detect all bots.

In our experiment, the average overlap values are calculated across 5 consecutive time windows with respect to the current time window. The length of the time window is set to be 10 minutes. We have tested with various window parameters and acquired similar detection results. However, choosing a relatively small window and short window sequence results in more prompt detection. Otherwise, it causes longer detection delay if more subsequent time windows are used to calculate the average overlaps. In our parameter setting, the P2P C&C channels can be detected within 5 consecutive time windows. To demonstrate the effectiveness of the self-correlation scheme, the moving average overlaps for 10 initial time windows for each P2P host in the trace are shown in Figure 5. For each trace, the result for only one P2P host is depicted in the figure. Other hosts in the same trace all demonstrate similar patterns as the one reported here.

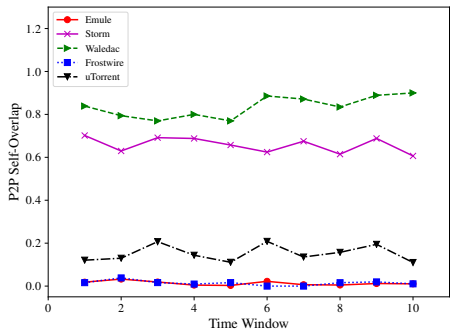


Fig. 5: Peer overlaps for botnet/normal P2P applications.

From Figure 5, there is significant difference between the

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

peer overlap for P2P bots and normal P2P hosts. In real world detection, a simple threshold based measure (e.g., 0.6) would suffice to distinguish between them. Based on the reported overlaps, our *p2p_c&c_detector* manages to identify all P2P bots in both traces. This demonstrates the effectiveness of the self-correlation scheme for detecting P2P C&C.

IRC C&C detection. In the Rbot botnet, the bot communicates with the C&C server through an established IRC channel. Via this channel the server commands the bot to perform port scanning against certain IPs in the network and the bot continuously reports scanning results to the C&C server. Our *irc_c&c_detector* can accurately detect the IRC C&C channel using keyword matching with negligible delay since the detection is performed entirely in kernel space.

Different from other botnets, Menti adopts a custom protocol and performs port scans. Since its C&C communication does not rely on the three well-known C&C protocols, *c&c_detector* discovers no suspicious C&C pattern. However, our experiment shows that *online_ml_detector* raises an alert since suspicious activities are discovered in the trace, which will be further explained in the following analysis.

Runtime ML based bot detection. As shown in Table III, the suspicious ratios for the majority bots fall between 95% to 99%. For the remaining bots, the ratios still exceed 90%. The result for Waledac is relatively low since the majority of its traffic is C&C related. Since the goal of ML detection module is to detect suspicious bots instead of individual connections, *online_ml_detector* raises alerts based on the suspicious ratio for each individual host. As defined in Section II, it represents the ratio of the number of suspicious connections with respect to the total number of connections for each host. If the ratio exceeds a pre-specified threshold, the host will be included in a greylist. Obviously, the choice of this threshold has a direct impact on the detection accuracy. A higher threshold leads to more false negatives while a lower threshold incurs more false positives. With a threshold of 10%, in the normal trace, 4 out of 36 are false positives. Instead, with a higher threshold, for example, 85%, all bots are accurately identified, with zero false positives. By excluding hosts with a limited number of connections, the ratios of all other hosts are shown in Figure 6.

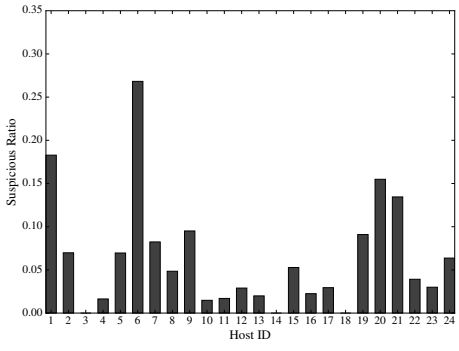


Fig. 6: Suspicious ratios for hosts in the normal trace.

For novel botnets, such as Menti, no C&C patterns are discovered since it uses a custom C&C protocol. However, as shown in Table III the *online_ml_detector* reports high suspicious ratio for the bot in this trace and raises alarms for further correlation. This parallel design is extremely effective in detecting bots in real-world, since it increases the difficulty of novel botnet variants evading both stages of detection.

Performance comparisons. We compare our approach to previous works by evaluating them using the same datasets in our experiment. Since our detection relies on parallel correlation of C&C detection and bot activity detection, the comparison is two-fold. First, we compare our ML detection accuracy to a recent ML based approach [13]. Different from ours, their ML model is based on per-flow feature vectors and only achieves an overall accuracy of 75% on a testing dataset containing multiple novel botnet variants not embraced in the training dataset. By contrast, in our evaluation for Menti (a variant with a custom C&C protocol), the ML model reports a fairly high accuracy (∼99.4%). Indeed our detection method achieves satisfactory accuracy for a majority of the test traces.

Moreover, another similar work [11] focuses on the detection of stealthy P2P botnets through cross-bot overlap analysis, which shows that P2P bots could be identified based on cross-bot correlation. Our experiments evaluate the detection accuracy on the same P2P botnets. As shown in Table III, all P2P bots are detected without any false negatives, which indicates that Botsifter can achieve comparable detection performance by solely using single-bot patterns. However, our approach is more robust in the scenario of a small number of bots.

V. Related Work

SDN based detection The emerging SDN techniques offer new opportunities for enhancing network security. The visibility and programmability provided by SDN have often been leveraged to perform various security detection and attack mitigation schemes.

FRESCO [25] represents one of the initial attempts to compose security services for SDN network systems. It is a framework to enable rapid development of OpenFlow-based security applications on the control plane. Shin *et al.* implemented a FRESCO version of the BotMiner [9] to perform clustering and correlations over network traffic to identify bot hosts. Later, Sonchack *et al.* proposed OFX [26], a system that enables practical deployment of security functions within an existing OpenFlow infrastructure. It allows control applications to dynamically load security modules directly into unmodified SDN compatible switches. OFX integrates botnet detection as a sample security application. It loads pre-processing functions into the switches so that the switches could send batch updates containing the processed feature vectors to the controller for further processing. However, neither solution utilizes computational power of the SDN data plane to perform more fine-grained and accurate detection.

Machine learning based detection In the early stage of bot detection, numerous efforts focused on identifying unique behavioral patterns of bots. For this purpose, various techniques are employed, such as matching the IDS dialog [8] and statistical algorithms to detect organized network activities [10].

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

However, a big challenge faced by these solutions is the prior knowledge of the malicious behaviors. Alternatively, the advancement of machine learning techniques and frameworks facilitate the explorations of the feature space of malicious traffic to address the challenge.

BotFinder [27] creates a model based on statistical features of flows to detect C&C communications. The model is built on a clustering algorithm to capture similar malicious behaviors of flows. Later, Zhao *et al.* proposed a detection method that employs a decision tree based algorithm with feature vectors extracted from the flows using a sliding window technique. In this model, the feature space is expanded to 12 features of flows [12]. Nonetheless, both algorithms are created towards specific botnet families or certain types of botnet, e.g., P2P botnet. Furthermore, most existing approaches are offline since real-time detection with a rich feature set is expensive and may impact network performance significantly.

Targeting the newly emerged cloud hosted bots, BotSifter addresses all these limitations by utilizing distributed detection and centralized learning in SDN, enhanced with a scalable traffic monitoring facility, and thus offering better detection efficiency.

VI. Conclusion

The cloud-based bots pose an imperative threat to the applications and services running on various cloud platforms, yet highly accurate and scalable detection solutions are not available. In this study, we have designed and implemented BotSifter, an SDN based scalable and accurate runtime bot detection framework in data centers. BotSifter utilizes a centralized learning and distributed detection model that is effectively supported by SDN. Furthermore, BotSifter adopts parallel detection of both the botnet C&C communication patterns and the machine learning based attack activities. The evaluations show the effectiveness of BotSifter based on real-world traces.

VII. Acknowledgement

We appreciate the constructive comments from the reviewers. This work is supported by NIST grant 70NANB18H272 and NSF grant CNS-1524462. This work was also supported in part by the Institute for Smart, Secure and Connected Systems at Case Western Reserve University through a grant provided by the Cleveland Foundation.

References

[1] L. Columbus, "83% Of Enterprise Workloads Will Be In The Cloud By 2020," 2018. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2018/01/07/83-of-enterprise-workloads-will-be-in-the-cloud-by-2020/#32e7847e6261

[2] P. McDougall, "Microsoft: Kelihos Ring Sold 'Botnet-As-A-Service'," 2011-09-30. [Online]. Available: https://www.darkreading.com/risk-management/microsoft-kelihos-ring-sold-botnet-as-a-service/d/d-id/1100470?piddl_msgorder=thrd

[3] K. Clark, M. Warnier, and F. M. Brazier, "Botclouds-the future of cloud-based botnets," in *in CLOSER*. Citeseer, 2011.

[4] [Online]. Available: https://securelist.com/elasticsearch-vuln-abuse-on-amazon-cloud-and-more-for-ddos-and-profit/65192/

[5] [Online]. Available: https://www.elastic.co/

[6] B. Butler, "Hackers found controlling malware and botnets from the cloud," 2014-06-26. [Online]. Available: https://www.networkworld.com/article/2369887/cloud-security/hackers-found-controlling-malware-and-botnets-from-the-cloud.html

[7] J. Goebel and T. Holz, "Rishi: Identify bot contaminated hosts by irc nickname evaluation." *HotBots*, vol. 7, pp. 8–8, 2007.

[8] G. Gu, P. A. Porras, V. Yegneswaran, M. W. Fong, and W. Lee, "Bothunter: Detecting malware infection through ids-driven dialog correlation." in *USENIX Security Symposium*, vol. 7, 2007, pp. 1–16.

[9] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection," 2008.

[10] G. Gu, J. Zhang, and W. Lee, "Botsniffer: Detecting botnet command and control channels in network traffic," 2008.

[11] J. Zhang, R. Perdisci, W. Lee, U. Sarfraz, and X. Luo, "Detecting stealthy p2p botnets using statistical traffic fingerprints," in *Dependable Systems & Networks (DSN), 2011 IEEE/IFIP 41st International Conference on*. IEEE, 2011, pp. 121–132.

[12] D. Zhao, I. Traore, B. Sayed, W. Lu, S. Saad, A. Ghorbani, and D. Garant, "Botnet detection based on traffic behavior analysis and flow intervals," *Computers & Security*, vol. 39, pp. 2–16, 2013.

[13] E. B. Beigi, H. H. Jazi, N. Stakhanova, and A. A. Ghorbani, "Towards effective feature selection in machine learning-based botnet detection approaches," in *Communications and Network Security (CNS), 2014 IEEE Conference on*. IEEE, 2014, pp. 247–255.

[14] T. Holz, M. Steiner, F. Dahl, E. Biersack, F. C. Freiling *et al.*, "Measurements and mitigation of peer-to-peer-based botnets: A case study on storm worm." *LEET*, vol. 8, no. 1, pp. 1–9, 2008.

[15] B. B. Kang, E. Chan-Tin, C. P. Lee, J. Tyra, H. J. Kang, C. Nunnery, Z. Wadler, G. Sinclair, N. Hopper, D. Dagon *et al.*, "Towards complete node enumeration in a peer-to-peer botnet," in *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security*. ACM, 2009, pp. 23–34.

[16] S. Stover, D. Dittrich, J. Hernandez, and S. Dietrich, "Analysis of the storm and nugache trojans: P2p is here," *USENIX; login*, vol. 32, no. 6, pp. 18–27, 2007.

[17] G. K. Venkatesh and R. A. Nadarajan, "Http botnet detection using adaptive learning rate multilayer feed-forward neural network," in *IFIP International Workshop on Information Security Theory and Practice*. Springer, 2012, pp. 38–48.

[18] T. Cai and F. Zou, "Detecting http botnet with clustering network traffic," in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on*. IEEE, 2012, pp. 1–7.

[19] L. Carl *et al.*, "Using machine learning technliques to identify botnet traffic," in *Local Computer Networks, Proceedings 2006 31st IEEE Conference on. IEEE*, 2006.

[20] S. Garcıa, "Identifying, modeling and detecting botnet behaviors in the network," *Unpublished doctoral dissertation, Universidad Nacional del Centro de la Provincia de Buenos Aires*, 2014.

[21] [Online]. Available: https://www.tensorflow.org/

[22] S. Garcia, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *computers & security*, vol. 45, pp. 100–123, 2014.

[23] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *computers & security*, vol. 31, no. 3, pp. 357–374, 2012.

[24] B. Rahbarinia, R. Perdisci, A. Lanzi, and K. Li, "Peerrush: Mining for unwanted p2p traffic," in *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 2013, pp. 62–82.

[25] S. W. Shin, P. Porras, V. Yegneswara, M. Fong, G. Gu, and M. Tyson, "Fresco: Modular composable security services for software-defined networks," in *20th Annual Network & Distributed System Security Symposium*. NDSS, 2013.

[26] J. Sonchack, J. M. Smith, A. J. Aviv, and E. Keller, "Enabling practical software-defined networking security applications with ofx." in *NDSS*, 2016.

[27] F. Tegeler, X. Fu, G. Vigna, and C. Kruegel, "Botfinder: Finding bots in network traffic without deep packet inspection," in *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, 2012.

Wang, An; Zha, Zili; Guo, Yang; Montgomery, Douglas; Chen, Songqing.
"BotSifter: A SDN-based Online Bot Detection Framework in Data Centers."
Paper presented at IEEE CNS 2019 - 2019 IEEE Conference on Communications and Network Security (CNS), Washington, D.C., United States.
June 10, 2019 - June 12, 2019.

MATCHING AND COMPARING OBJECTS IN A SERIAL CYTOMETER

Nikita Podobedov[1,2], Matthew DiSalvo[2,3], Jason Hsu[2,4], Paul Patrone[2], and Gregory A. Cooksey[2]

[1]Columbia University, [2]National Institute of Standards and Technology (NIST),
[3]Johns Hopkins University, [4]Montgomery Blair High School; USA

ABSTRACT

Flow cytometers are indispensable for clinical studies yet are limited by inherent uncertainties. We have developed an optofluidic device capable of multiple measurements along a microfluidic channel, whereby many of the uncertainty components can be systematically evaluated and reduced. This study addresses the challenge associated with identifying and tracking the order of objects throughout their time of transit. An algorithm was developed to characterize objects as they travel. We discuss methods of testing the efficacy of this algorithm, and other tools that allow us to gain more information than is provided by a conventional cytometer.

KEYWORDS: Flow Cytometry, Microfluidics, Optofluidics, Reproducibility

INTRODUCTION

Traditional flow cytometers can measure thousands of cells per second, but the intensity of the fluorescent biomarkers on cells is only measured once and used to calculate scalar values such as integrated area.[1] Thus, cytometers have limited ability to characterize biomarker distributions with high precision, which reduces their capacity to discriminate small changes within a population and to distinguish rare objects. Our group is developing an optofluidic cytometer that increases measurement resolution and precision by repeating measurements four times along the flow path: two axially symmetric waveguides collect fluorescence at each of two measurement regions separated along a microchannel. In order to aggregate the replicate measurements, an algorithm was created in MATLAB (MathWorks) to perform the piecewise temporal alignment across all four data channels. The algorithm evaluates the likelihood of object ordering based on time-of-flight and shape of the fluorescence signal, which includes high-temporal resolution of the intensity as the particle moves through the measurement region. We discuss challenges associated with objects passing one another, scenarios with multiple particles in the measurement region, such as doublets, and the unique opportunities that our high-resolution approach provides in extracting detailed information from objects, which would normally be impossible with other approaches. We also discuss metrics to assess the efficacy of matching and doublet separation.

EXPERIMENTAL

(DISCLAIMER: Identification of commercial products does not imply recommendation or endorsement by NIST. The materials and equipment used may not necessarily be best for purpose.) The manufacturing process for these devices has been detailed in a previous publication.[2] Briefly, two poly(dimethylsiloxane) (PDMS) layers with lithographed microchannels were aligned and bonded. The two measurement regions each included one waveguide which transmitted laser light and two symmetrically angled waveguides that collected emitted fluorescence from objects (15 μm diameter microspheres, Bangs Lab FSDG009) passing the laser (Fig. 1A). Fluorescence emission was measured on a photomultiplier tube (Hamamatsu H11903-20) and digitized (National Instruments PCIe-6374).

RESULTS AND DISCUSSION

Fluorescent microspheres passing through two measurement regions (each with a fluorescence collector angled upstream and downstream of the excitation) created signals as shown in Fig. 1. Calibration of our serial cytometer requires that standard objects be compared between regions to establish a scale factor for normalization. Importantly, this exercise requires unambiguous object matching from one region to another. Object matching was strongly dependent on a consistent particle velocity, which we maintained to within 0.2 % coefficient of variation (CV) using a combination of inertial and hydrodynamic focusing. Coincident events (doublets) impose additional challenges of identification and decoupling. Because our system records an object's intensity both in time and from different perspectives at two different regions (Fig. 1B), a number of methods were tested in MATLAB, including idealized peak fitting, to facilitate individual object resolution and ordering (Fig. 1D, E).

To test algorithm efficiency, a number of *in silico* simulations were performed. Experimental data were lowpass filtered, normalized, temporally aligned, and averaged to form an idealized signal. We then modeled distributions of variables such as particle velocity, time between particles, white noise, and particle intensities from these data.

Cooksey, Gregory; Patrone, Paul; Hands, James; Meek, Stephen; Kearsley, Anthony.
"Matching and Comparing Objects in a Serial Cytometer."
Paper presented at 22nd International Conference on Miniaturized Systems for Chemistry and Life Sciences (MicroTAS 2018), Kaohsiung, Taiwan Province of China. November 11, 2018 - November 15, 2018.

Lastly, idealized signals were scaled, time-shifted, and had noise added to them to match the modeled distributions of those components, thereby creating simulated data equivalent to that of a mock bead population. Overall, when analyzing artificial signals, our algorithms correctly determined the identity of synthetic events across replicate measurement channels with 100 % tracking efficiency. In extracting integrated fluorescence area from the artificial signals, our algorithms demonstrated a precision of ± 0.04 % (± 1 standard deviation) compared to the synthetic ground truth. Using hydrodynamic focusing to position particles in the center of the channel, with event rates near what is typically achieved by commercial cytometers (1000 events per second), over 10 % of objects were predicted to pass each other between regions and ≈ 25 % could manifest as doublets (Fig. 1F). Adjusting hydrodynamic focusing to align particles to a single inertial focusing node reduced the velocity distribution to 0.2 % CV, leading to a 10-fold improvement in passing probability and 2-fold improvement in doublet likelihood.

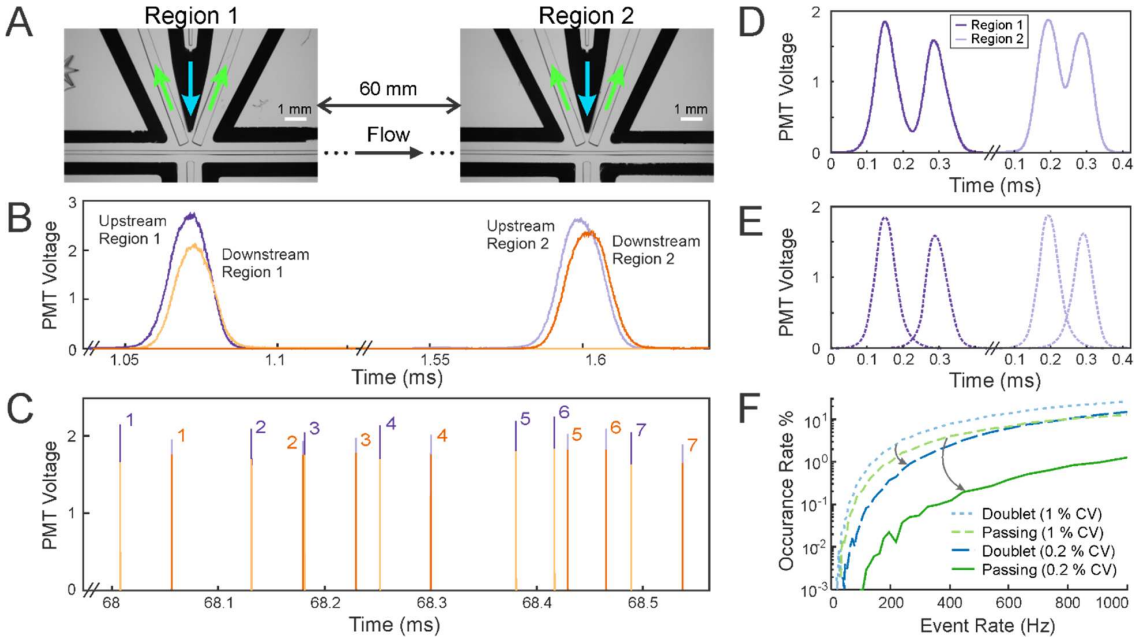


Figure 1: (A) Microscopy images of the 2 measurement regions in the microfluidic flow cytometer. Cyan: laser excitation; Green: fluorescence emission. (B) Representative fluorescence peaks as a bead travels through the 2 measurement regions. (C) Representative bead matching from time-of-flight and shape analysis. Numbers show tracking of each bead's data at region 1 (purple) and region 2 (orange). (D) Example of measured doublet and (E) simulated separation of each bead from the signal in both regions. (F) Simulation results comparing average rates of objects entering the device and the occurrence of passing and doublet events before (1 % velocity CV) and after (0.2 % velocity CV) improvement in flow focusing using combination of hydrodynamic and inertial focusing.

CONCLUSION

By providing multiple high-resolution measurements of objects in flow, our serial flow cytometer enables characterization of particles while promising improved discrimination of aberrant signal events. Importantly, we have demonstrated the need for an effective tracking and fitting algorithm for any microfluidic system that implements repeated measurement of objects along a flow path. Work is ongoing to extract additional measurement features from the upstream and downstream projections on the same object, including how these measurements can be used to extract the particle's size, shape, and fluorophore distribution.

REFERENCES

[1] H.M. Shapiro, "Practical Flow Cytometry, 4th edition," John Wiley & Sons Inc., New York, 2003.

[2] G.A. Cooksey, P.N. Patrone, et al., "Dynamic Measurement of Nanoflows: Realization of an Optofluidic Flow Meter to the Nanoliter-per-Minute Scale". *Anal. Chem.*, 91 (2019), pp. 10713-10722.

CONTACT

* G. Cooksey; phone: +1-301-975-5529; gregory.cooksey@nist.gov

Cooksey, Gregory; Patrone, Paul; Hands, James; Meek, Stephen; Kearsley, Anthony.
"Matching and Comparing Objects in a Serial Cytometer."
Paper presented at 22nd International Conference on Miniaturized Systems for Chemistry and Life Sciences (MicroTAS 2018), Kaohsiung, Taiwan Province of China. November 11, 2018 - November 15, 2018.

Anisotropy of dynamic disorder and mobility in single crystal tetracene transistors.

Emily G. Bittle(1), Adam J. Biacchi(1), Lisa A. Fredin(2), Andrew Herzing(3), Thomas C. Allison(2), Angela R. Hight Walker(1), David J. Gundlach(1)

(1) Nanoelectronics Group, Engineering Physics Division, PML, (2) Chemical Informatics Group, Chemical Sciences Division, MML, (3) Materials Structure and Data Group, Materials Measurement Science Division, MML

250 Word Abstract:

Recent interest on the topic of dynamic disorder in organic semiconductors and the influence on electrical transport has inspired several interesting experimental studies which evaluate the amount of dynamic disorder and relative device mobility in organic semiconductors. In this study, we investigate the directionality of intermolecular vibrational modes and the anisotropic mobility in tetracene single crystals to gain an understanding of the relationship between specific motions and electrical transport. We use transmission electron microscopy (TEM) and low-frequency polarized Raman to measure vibrational modes, and anisotropic mobility is evaluated using a field-effect transistor on the same crystal used for Raman study. To identify the modes measured and provide an estimate of the influence that the related motions are expected to have on the anisotropic transfer integrals, we use a combination of computational methods. We observe five low-frequency Raman-active modes in the ab plane of the crystal, and a streaking pattern is found in one direction using TEM indicative of motion in the crystal. Results relate the directionality of the mode motions and the highest and lowest mobility in the tetracene crystal.

100 Word Abstract:

This study investigates the influence of directionally oriented dynamic disorder on electrical transport in organic single crystal transistors. The intermolecular vibrational modes in tetracene single crystals are measured using transmission electron microscopy and low-frequency polarized Raman, and mobility is obtained from a field effect transistor geometry on the same crystal used for Raman study. A combination of computation methods are used to identify the modes measured and provide an estimate of the influence that the related motions are expected to have on the anisotropic transfer integrals. Results discuss the relationship between the directionality of these motions and the highest and lowest mobility in the tetracene crystal.

Bittle, Emily; Biacchi, Adam; Fredin, Lisa; Herzing, Andrew; Allison, Thomas; Hight Walker, Angela; Gundlach, David.
"Anisotropy of dynamic disorder and mobility in single crystal tetracene transistors.."
Paper presented at SPIE Optics and Photonics 2018, San Diego, CA, United States. August 19, 2018 - August 23, 2018.

Case Study – Exploring Children's Password Knowledge and Practices

Yee-Yin Choong, Mary Theofanos
National Institute of Standards and Technology
{yee-yin.choong,mary.theofanos}@nist.gov

Karen Renaud, Suzanne Prior
Abertay University, Dundee
{k.renaud,s.prior}@abertay.ac.uk

***Abstract*—Children use technology from a very young age, and often have to authenticate themselves. Yet very little attention has been paid to designing authentication specifically for this particular target group. The usual practice is to deploy the ubiquitous password, and this might well be a suboptimal choice. Designing authentication for children requires acknowledgement of child-specific developmental challenges related to literacy, cognitive abilities and differing developmental stages. Understanding the current state of play is essential, to deliver insights that can inform the development of child-centred authentication mechanisms and processes. We carried out a systematic literature review of all research related to children and authentication since 2000. A distinct research gap emerged from the analysis. Thus, we designed and administered a survey to school children in the United States (US), so as to gain insights into their current password usage and behaviors. This paper reports preliminary results from a case study of 189 children (part of a much larger research effort). The findings highlight age-related differences in children's password understanding and practices. We also discovered that children confuse concepts of *safety* and *security*. We conclude by suggesting directions for future research. This paper reports on work in progress.**

I. Introduction

Systems designed specifically for children are becoming increasingly popular. Many of these authenticate the child in order to retain a history of interaction, or to ensure that it is genuinely a child using the system. Usability testing with children is constrained by strict ethical requirements [59], [39], which might put researchers off testing alternative authentication mechanisms with this target group altogether. Without evidence of clearly superior and appropriate alter-natives, it is understandable that developers revert to the password. This might well be a suboptimal choice for this user group due to issues such as as heterogeneiety in ability [87], language proficiency [57] and immature literacy [51].

Most of the research in usable security has focused on adults. Yet over the next 10 to 20 years the world's cyber posture and culture will be dependent on the cybersecurity and privacy knowledge and practices of today's youth. We performed a systematic literature review which highlighted a lack of research examining extant child password use (Section II). Without an understanding of extant behavior, it is infeasible to start seeking an alternative, more appropriate, mechanism for child-tailored authentication. Thus, having identified this need, we proceeded to survey school children in the United States (US) to bridge this gap.

We reflect on our findings and suggest that we ought to prepare our children more effectively for a world where password hygiene is crucial. It is important to teach principles at a young age, and not haphazardly, which the evidence suggests is being done at present. We should deliberately nurture the development of good password hygiene habits *as and when* children are first learning to use passwords. We should introduce more advanced concepts as they develop and are able to adopt them. This would create a foundation for responsible adult password usage.

II. Systematic Literature Review

A rigorous literature search must be valid and reliable [94]. *Validity* is ensured by rigorously identifying: selected databases, keywords and inclusion and exclusion criteria; covered period; and applied forward and backward search. *Reliability* is based on the fully documented search process.

We searched through the following databases: Google Scholar[1], ACM, DBLP, IEEEXPlore, ScienceDirect and SpringerLink, http://worldcat.org PhD Theses, Educational literature: ERIC - Dept. of Education (Eric.ed.gov), Popular Press: Google News. We started off with Google Scholar and then augmented the list with articles from the other databases.

Key words: We used the following keywords for our search:

- Children/Minors/Teenagers/Adolescents and Authentication
- Children/Minors/Teenagers/Adolescents and Passwords
- Children/Minors/Teenagers/Adolescents and Security

Inclusion criteria: We chose to not only focus on high-quality literature so we also included lower rated journals, conferences and workshops as well as publications in the public media. We searched from the year 2000 onwards.

[1]Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

Workshop on Usable Security (USEC) 2019
24 February 2019, San Diego, CA, USA
ISBN 1-891562-57-6
https://dx.doi.org/10.14722/usec.2019.23027
www.ndss-symposium.org

Exclusion criteria: We filtered out publications that were not written in English and those published before the year 2000, as well as patents.

Forward search: We used Google Scholar citation index service to perform a forward search.

Backward search: The backward search was conducted manually.

We conducted a full-text search, and the databases were searched to identify publications that contained the keywords. We then applied the defined inclusion and exclusion criteria. Following this process only 87 publi-cations remained. Table I shows the number of publications for each database search, how many were excluded and how many remained for analysis. It should be noted that some publications appeared in more than one database.

TABLE I. OUTCOME OF LITERATURE SEARCH OF ALL DATABASES

Source	# papers	# eliminated	# retained
Google Scholar	65	13	52
ACM	2	0	2
DBLP	2	0	2
IEEEExplore	4	2	2
ScienceDirect	8	1	7
WorldCat	0	0	0
Eric	2	2	0
Google News	4	2	2
Scopus	4	1	3
Public Media	17	0	17

Table II summarizes the range of reports we analysed. Despite several of the retrieved papers considering new authentication methods, only four empirical studies were found which used children as participants in the development or evaluation of child-specific authentication [71], [16], [18], [54]. Of these, only Coggins [17] reported conducting a pre-study measure of the children's knowledge and experience of passwords. However, Coggins does not include information on what these pre-measures showed or what bearing they had on the final result.

This literature review revealed a gap in the literature related to gauging current levels of comprehension and practice related to passwords. Filling this gap is important because researchers conducting empirical studies in this area benefit from understanding the current level of password knowledge held by children and indeed some indication of their current password practices.

III. RESEARCH METHODOLOGY

Given the limited work to date in measuring children's knowledge and experience of passwords, there is a need for more studies to add to these findings in different contexts. In this study, we developed a self-report survey to understand what challenges children grades 3 through 12 face regarding passwords. The goal was to identify students' practices, perceptions, and knowledge regarding passwords. Each student answered questions assessing their use of computers, passwords, password practices, knowledge about and feelings about passwords, together with information about grade and gender. We wanted to address the following research questions (RQ):

TABLE II. CATEGORIES FROM ANALYSIS

Classification	Refs
Designing for Children	
Guidelines for keeping children safe online	[5], [13], [22], [25], [27], [35], [47], [64], [67], [73], [79], [80], [11], [37], [83], [24], [95], [27], [49], [50], [40], [84], [66], [6], [77], [3]
Argument for children's privacy rights	[8], [86]
Study of children's Internet Usage	[15], [32], [36], [45], [55], [82], [85], [44], [49], [23], [21], [48], [78], [43], [61], [62], [63], [16], [58], [69], [74], [81], [75], [76], [12]
Accessibility Issues	[29]
Security awareness	[2], [9], [31], [46], [69], [70], [91], [33], [88], [89], [93], [52], [17], [53], [65], [92]
Children Accessing Adult Content	[96], [72]
Children & Authentication	
Proposes new authentication ideas	[60], [38], [19]
Empirical Study of Passwords and Children	[42], [71], [54], [41]
Empirical Evaluation of Child-Specific Authentication Methods	[71], [16], [18], [54]
Designing Security for children	[30], [68], [26]
Anecdotal reports of children's password behaviors	[1], [10], [7], [20], [28]
Children & Biometrics	[4], [90]
Proposes new authentication ideas	[60], [38], [19]

RQ1. How do children currently use Computers and Passwords?

RQ2. Password Understanding:

(a) Password Hygiene Knowledge?
(b) Why do they need passwords?
(c) What are students' passwords perceptions?
(d) Do they know how to create a strong password?

RQ3. Password Behaviors:

(a) How do students select, remember and store passwords?
(b) What are the characteristics of the password they are asked to formulate to access a game?

A. Survey Development

The research questions guided the development of objectives for accessing student's use of computers, of passwords, password practices, knowledge about passwords, feelings about passwords, and tests for gender, age and school differences. A list of possible items was generated targeting the objectives, as illustrated in the alignment matrix in the Appendix.

Two surveys were designed: one 15 item survey for grades 3 to 5, and a 16 item survey for grades 6 to 12. All of the items were closed response except for four open response items where students were asked: how many passwords they have; how many times a day they use passwords; to list a reason(s) why people should use passwords, and to generate a new password for a given scenario. While the surveys were identical in item content, the language and format of the

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

response variables were tailored for appropriateness to the age groups. For example, most of the response variables were "Yes" or "No" for the 3^{rd} - 5^{th} graders, while the 6^{th} - 12th graders' response variables were lists of check all that apply. To ascertain the content and construct validity of the survey instruments, four types of reviews were conducted. Content experts in usable security were asked to evaluate the alignment matrix (in the Appendix) and provide feedback on the alignment of the categories with the scope of the survey goals, of the alignment of the items with the category, and if there were missing items. Survey experts also reviewed each item for clarity for the intended audience, appropriate format for what the item is assessing, and alignment of response options. Content experts (elementary, middle and high school teachers) focused on the language and format of the items based on the grade/age of the students. Cognitive interviews with students, to determine if the questions were indeed being interpreted as intended, were also conducted using a talk-aloud protocol. Cognitive probing techniques where students were asked to both paraphrase items (e.g., "How would you ask the question in your own words") and interpret them (e.g., "What is your answer and why") complemented the talk-aloud protocol. Additionally, we piloted the surveys with students. After each type of review, the survey instruments were refined based on the feedback and comments.

B. Procedure & Recruitment

The study was approved by the full Institutional Review Board. Principals and teachers were recruited to participate. The schools, individual teachers, and students that participated were compensated. Each school received $1000, the teachers received $50 gift cards, and the students received age appropriate trinkets such as caricature erasers or ear buds, for example. Finally, parental consent and student assent forms were collected prior to survey distribution. Students who did not receive parental consent performed alternative activities following the school's standard protocol during the survey administration. Each participating classroom also received $50 for a classroom thank-you celebration where all students celebrated, including those who did not participate in the survey. The survey administration was tailored for the appropriate age group. All children completed scantron survey forms, with teachers reading the survey aloud in the 3^{rd} - 5^{th} grades.

The results presented here are the initial results from only two US Midwest schools - an elementary (3^{rd} - 5^{th} grades) and a middle school (6^{th} - 8^{th} grades). This is the first data set from a much larger research effort that will include between 1500 to 1800 elementary, middle and high school students from up to six US school districts.

C. Participants

In this case study dataset, a total of 189 school students completed the surveys. Both schools are located in the Midwest region in the US. Table III presents the participant demographics.

TABLE III. PARTICIPANT DEMOGRAPHICS

Graders	**#**	**M:Male F:Female U:Unspec**	**Age Years**	$\bar{x}$	σ
3^{rd} - 5^{th}	88	M:53.41% F:42.05% U:4.54%	8-12	8.15	0.96
6^{th} - 8^{th}	101	M:51.49% F:40.59% U:7.92%	11-15	12.55	1.03

D. RQ1:Current Usage

1) Current Computer Usage: The most popular type of computers the 3^{rd} - 5^{th} graders use was gaming console (80.68%), followed by laptop (78.41%) and tablet (71.59%). For the 6^{th} - 8^{th} graders, the most popular type of computers was cell phone (87.13%), followed by laptop (80.20%) and gaming console (77.23%). Table IV is a list of all types of computers used by the participants in the survey.

TABLE IV. USAGE OF DIFFERENTS KINDS OF COMPUTERS

	Desktop	**Laptop**	**Tablet**	**Cell Phone**	**Gaming Console**
3^{rd} - 5^{th}	62.5%	78.41%	71.59%	68.18%	80.68%
6^{th} - 8^{th}	64.36%	80.20%	55.45%	87.13%	77.23%

Locations where students use the computers were mostly at school or at home: the 3^{rd} - 5^{th} graders reported computer use at school (98.86%) and at home (81.82%); the 6^{th} - 8^{th} graders reported similar computer use at school (94.06%) and a higher usage at home (91.09%).

Activities that students use computers for ranged from school work, homework to games and social media. Games (92.05% for 3^{rd} - 5^{th} and 84.16% for 6^{th} - 8^{th}) and entertainment (89.77% for 3^{rd} - 5^{th} and 85.15% for 6^{th} - 8^{th}) were the most popular activities on computers for both groups. Table V is a list of computer activities sorted by 3^{rd} - 5^{th} graders' percentages. Percentages of 6^{th} - 8^{th} graders follow a similar pattern.

TABLE V. ACTIVITIES USING COMPUTERS

	3^{rd} - 5^{th}	**6^{th} - 8^{th}**
Games	92.05%	84.16%
Entertainment	89.77%	85.15%
Internet	80.68%	77.23%
School	75.00%	78.22%
Texting	45.45%	56.44%
Social media	43.18%	63.37%
Email	36.36%	35.64%
Homework	32.95%	55.45%

2) Current Password Usage: On average (medians), the 3^{rd} - 5^{th} graders have 2 passwords at school and 3 passwords at home. They reported to use their passwords about 4 times a day. For the 6^{th} - 8^{th} graders, they have on average (medians) 2 passwords at school and 4 passwords at home; use passwords about 4 times a day. More than 90% in both groups reported that they use passwords to access school computers, and between about 71% to 80% reported using passwords to unlock home computers. Table VI lists technologies locked with passwords reported by participants in the survey.

E. RQ2: Understanding of Password Hygiene

1) Knowledge: The 3^{rd} - 5^{th} graders reported learning about good password use more at home (71.59%) compared

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

TABLE VI. WHAT CHILDREN AUTHENTICATE TO USE

	3^{rd} - 5^{th}	6^{th} - 8^{th}
School computers	98.86%	94.06%
Home computers	71.59%	80.20%
Tablets	60.23%	55.45%
Cell phones	62.50%	85.15%
Games	55.68%	40.59%
Email	29.55%	45.54%
Social media	45.45%	69.31%

to at school (38.64%); whereas the 6^{th} 8^{th} graders reported learning with almost equal percentages at school (73.27%) and at home (76.24%).

Regarding what they know about password hygiene, responses of "Always" and "Sometimes" were combined, shown in Table VII. More than 90% of each age group reported that they keep their passwords private; and they keep a good habit of signing out after computer use (89.77% for the 3^{rd} - 5^{th} graders and 94.06% for the 6^{th} - 8^{th} graders). The 6^{th} - 8^{th} graders tend to share their passwords with friends more.

TABLE VII. KNOWLEDGE OF PASSWORD HYGIENE

	3^{rd} - 5^{th}	6^{th} - 8^{th}
Keep passwords private	90.91%	93.70%
Sign out after computer use	89.77%	94.06%
Share with friends	32.95%	47.52%
Use same password for everything	57.95%	78.22%
Change passwords	62.50%	79.21%

For those who reported that they changed their passwords, the top reasons were "when someone finds out my passwords" (94.55% for the 3^{rd} - 5^{th} graders and 72.50% for the 6^{th} - 8^{th} graders) and "when I forgot my passwords" (54.55% for the 3^{rd} - 5^{th} graders and 68.75% for the 6^{th} - 8^{th} graders).

2) Why Passwords?: Both groups of participants were asked why it is important to use passwords. The 3^{rd} - 5^{th} graders were only asked to provide one reason while the 6^{th} - 8^{th} graders were asked to provide up to three reasons.

These responses were independently analysed by two authors using thematic coding. The authors met and agreed on the final codes. The agreed codes were at a very high level, reflecting the relatively concise answers given by the children. For example: "*To keep things safe*", "*Safety*", and "*People should use passwords so they can have privacy*". The final codes are shown in Table VIII; codes with fewer than 5 responses are not included. The higher number of responses attributed to codes within the 6^{th} 8^{th} graders are due to a higher number of participants and this group being asked to provide three responses, as opposed to one.

TABLE VIII. CODING OF SURVEY RESPONSES

3^{rd} - 5^{th} (n, % of responses)	6^{th} - 8^{th} (n, % of responses)
Privacy (34, 40.48%)	Privacy (81, 38.02%)
Safety (18, 21.42%)	Safety (41, 19.24%)
Access (15, 17.85%)	Security (40, 18.77%)
Security (9, 10.71%)	Access (24, 11.27%)
Hacking (6, 7.14%)	Hacking (10, 4.69%)

3) Password-Related Perceptions: The 3^{rd} - 5^{th} graders reported higher percentages of perceiving creating and remembering passwords as easy, than the 6^{th} - 8^{th} graders, as shown in Table IX. Both age groups found it fairly easy to enter passwords with a keyboard or using a touch screen. Although having too many passwords does not seem to be bothersome to either group, we do see a rising trend with older children having more passwords as they get older.

TABLE IX. PASSWORD PERCEPTIONS

	3^{rd} - 5^{th}	6^{th} - 8^{th}
Easy to make my password	76.14%	54.46%
Easy to make many different passwords	61.36%	44.55%
Easy to remember my passwords	80.68%	68.32%
Easy to enter my passwords with a keyboard	77.27%	80.20%
Easy to enter my passwords on a touch screen	71.59%	81.19%
I wish there was another way besides passwords	50.00%	31.68%
I have too many passwords	27.27%	16.83%

F. RQ3: Password Behaviors

1) Password Selection & Storage: When asked about how they get their passwords, about 85% in both age groups reported getting some passwords from schools.

Younger students (3^{rd} - 5^{th} graders) reported a high percentage of parental involvement in creating their passwords (either created by parents or they created their own passwords with help from parents, combined: 69.32%). A high percentage of the 6^{th} - 8^{th} graders (86.14%) reported creating their own passwords and low parental involvement (either created by parents or created their own passwords with help from parents, combined: 35.64%).

Participants reported memorizing their passwords (97.73% for the 3^{rd} - 5^{th} graders and 91.08% for the 6^{th} - 8^{th} graders). About a third of students in each group reported that they write passwords on paper. The 3^{rd} - 5^{th} graders also reported higher percentages relying on external sources (such as auto-fill by computer, family members remember for me, or save in a file on computer) compared to the 6^{th} - 8^{th} graders. Table X is a list of mechanisms of how students remember passwords.

TABLE X. HOW CHILDREN RETAIN PASSWORDS

	3^{rd} - 5^{th}		6^{th} - 8^{th}
	Always[2]	**Sometimes**	**Check all Apply**
Memorize	78.41%	19.32%	91.09%
Let computer save the password and auto-fill	32.95%	31.82%	36.63%
Write passwords down on paper	12.50%	23.86%	33.66%
Family member remembers for me	6.82%	28.41%	9.90%
Friend remembers for me	2.27%	7.95%	1.98%
Save in a file on computer	10.23%	87.50%	12.87%

The 6^{th} - 8^{th} graders were asked an additional question on whether they help their family members with passwords. Forty-Nine students (48.51%) reported "Yes," - of those, 73.47% reported helping family members to remember their passwords.

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

2) Created Password Analysis: The two groups were asked: "*Let's say you just got a new game to play on the computer, but you need a password to use it. Please make up a new password for that game. (Remember, don't write down one of your real passwords.)*".

G. Password Characteristics

The average (medians) lengths of the passwords created by participants were: 7 characters for the 3^{rd} - 5^{th} graders, with a range of [3, 32]; and 10 characters for the 6^{th} - 8^{th} graders, with a range of [4, 29]. Lowercase letters make up the majority of the passwords, followed by numbers. The most popular characters used were lowercase letters "a," "e," and "o" for the 3^{rd} - 5^{th} graders and "e," "a," and "r" for the 6^{th} - 8^{th} graders. The most used numbers for both age groups were "1" and "2." Symbols or white spaces were rarely used. Figure 1 shows the distribution of different character types used in the passwords created by the participants.

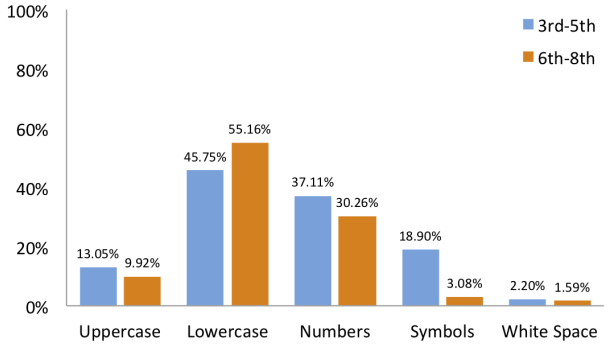


Fig. 1. Character Types in Passwords

We further examined character type positioning in the passwords. Figures 2 and 3 display the overall character type distribution relative to their position, for password lengths of 7 (for the 3^{rd} - 5^{th} graders) and password lengths of 10 (for the 6^{th} - 8^{th} graders).

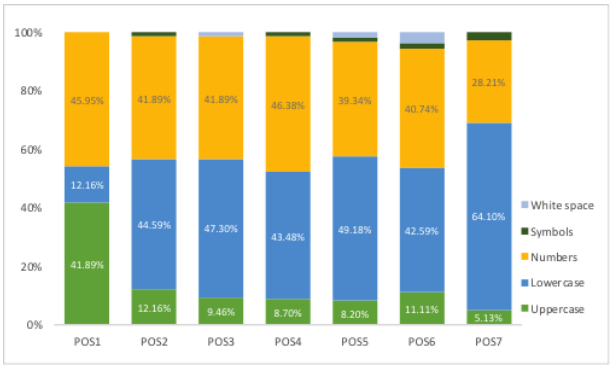


Fig. 2. Character Types by Positions in Passwords (3^{rd} - 5^{th})

As shown in Figure 2, in general, the percentages of the 3^{rd} - 5^{th} students who used lowercase letters and numbers were very close (around 40%) across all positions except for the two ends, i.e. position 1 (POS1) and position 7 (POS7). In POS1, 41.89% of the 3^{rd} - 5^{th} graders used uppercase letters in their passwords. In position 7, a lot more 3^{rd} - 5^{th} graders (64.10%) included lowercase letters in their passwords.

Fig. 3. Character Types by Positions in Passwords (6^{th} - 8^{th})

In contrast, the pattern for the 6^{th} - 8^{th} graders (Figure 3 looks quite different from that for the 3^{rd} - 5^{th} graders. A lot more 6^{th} - 8^{th} graders (between 52% and 64%) used lowercase letters across all positions except for the first position. Much fewer 6^{th} - 8^{th} students (between 20% and 39%) used numbers in their passwords compared to their younger counterparts. Similar to the 3^{rd} - 5^{th} graders, in POS1, 47.87% the 6^{th} - 8^{th} graders included uppercase letters in their passwords.

H. Password Strength

We used the `zxcvbn.js` JavaScript Library[3] to quantify the strength of the passwords the children formulated. Figure 4 shows the strengths of the two groups' passwords. There is an improvement as the children move up in the educational system, but the percentage of very weak passwords is still very high, even amongst the older group.

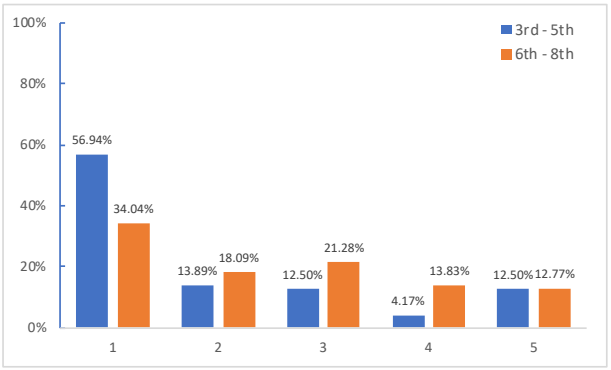


Fig. 4. Password Strengths

IV. DISCUSSION

A. RQ1: Password & Computer Usage

Not surprisingly, as children age, their use of technology and on-line activities change. The percentages of students

[3]https://www.bennish.net/password-strength-checker/

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

having cell phones increases almost 20% from the younger to the older children, as shown in Table IV. As the children mature, there is also about a 10% increase in some social activities, with texting increasing, and about a 20% increase in social media (in Table V). As a result, the older children experience more and more need for authentication: about 23% increase in cell phone authentication, 15% increase in email authentication, and 24% increase in social media authentication (Table VI). The increased need for authentication for older children translates into their having more passwords, each of which has to be retained and managed.

B. RQ2: Password Knowledge

Children demonstrated confusion between the concepts of passwords, privacy and safety or protection. Many children reported believing that passwords would keep them 'safe': "*I think that people should use passwords because it saves everyone's lives*" (Female, 3^{rd} 5^{th} grader)

Similar feelings were expressed in 22% of the younger group's responses and 19% of the elder group. This suggests that the message that passwords are required for security is becoming confused either in the delivery or in the children's receiving of the message with the idea of safety, forming inaccurate mental models. It is not clear why this is happening but scaring children into using passwords is unlikely to be the most effective method for ensuring they continue to develop good cyber security practices as they age. Resources in the area of cyber security for children are often focused on cyber bullying and the dangers of online predators, which may explain where some of the confusion arises from. Although educators and parents may not be intentionally trying to scare children into using passwords, children are receiving mixed messages. This demonstrates a need for clear and consistent messages. In order to achieve this, it may be necessary to provide additional training to those delivering the message, in this case parents and teachers.

C. RQ3: Password Practices and Behaviors

Children's ages influence their password practices and behaviors. Younger children rely more on their family in creating and remembering passwords. Almost twice as many of the younger group reported having parental help in creating their passwords. Moreover, about 35% of the younger children reported getting help from family members in remembering their pass-words, as compared to only 10% of the older children.

Parents also play an important role of providing guidance on 'good' password hygiene to the younger group. In contrast, schools play a larger role in influencing password behaviors of the older group. Moreover, the older children assist family members with remembering passwords. Both age groups understand that passwords should remain private and they sign out after computer use. However, approximately 50% of the older group reported sharing passwords with friends.

V. Reflection & Recommendations

These findings raise several questions. How do we best prepare children for the increasing demands of living with passwords? How do we provide guidelines and strategies as they age to help them learn and develop appropriate skills?

A. Password Management Lifecycle

It is necessary to consider users' password behaviors in a holistic manner, realizing that there are three stages in the password management lifecycle: password creation, password maintenance, and authentication. Users' behaviors are reflections of the interactions among stages in the lifecycle, the capabilities and limitations of the human information processor, and the individual factors [14]. Each stage requires cognitive capabilities from the password owner, in this case the child, to create, maintain, and authenticate using passwords. Overall, the 6^{th} - 8^{th} graders reported experiencing more difficulties with their passwords. Approximately 20% more of the 6^{th} - 8^{th} graders than the 3^{rd} - 5^{th} graders reported difficulties in creating passwords (Table IX). They also struggled more to maintain their passwords: 20% more of the older group reported using "same password for everything" (Table VII), probably because about 12% more of the older children found it difficult to remember passwords (Table X). It is important not to consider any of the lifecycle stages in isolation when we design authentication mechanisms for a particular target group, such as children.

B. Password Choice

On average, the 6^{th} - 8^{th} graders created passwords that were 3 characters longer than the 3^{rd} - 5^{th} graders did. Compared to the 6^{th} - 8^{th} graders, the 3^{rd} - 5^{th} graders used more uppercase letters, numbers, and white spaces when composing their passwords. The 3^{rd} - 5^{th} graders tend to start their passwords with numbers or uppercase letters in the 1st position. Immediately after the 1st position, lowercase letters and numbers dominate the next few positions until towards the end of the password where almost 2/3 of the characters are lowercase letters. The 6^{th} - 8^{th} graders also tend to start their passwords with uppercase letters, but numbers are not dominating as in the case of their younger counterparts. Immediately after the 1st position, the 6^{th} - 8^{th} graders use much higher percentage of lowercase letters than any other character types in all positions. Towards the end positions in the password, we observe slight rising trends of using numbers and symbols.

In a password generation study with 81 adults [56], the researchers found that uppercase letters dominate the 1st position in the password, then the rate of uppercase letters sharply drops at the 2nd position, while the lowercase letters substantially rise at position 2. Numbers follow a steady increasing trend and start dominating towards the latter positions in the password, until the last position where symbols make up half of the character distribution. The 6^{th} - 8^{th} graders' password trend resembles patterns of uppercase, lowercase, numbers, and symbols found in passwords generated by adults. This could be due to the fact that as students get older, they have more exposure to password complexity requirements for numbers and symbols.

The passwords that the participants created did not use a broad range of characters. For the 3^{rd} - 5^{th} graders, only 8 characters appeared with frequency higher than or equal to 3%, namely, 2, 1, a, e, o, 3, 5, and 6. For the 6^{th} - 8^{th} graders, only 11 characters appeared with frequency higher than or equal to 3%, namely, 2, e, 1, a, r, o, 4, n, l (lowercase), i and

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

s. Special character use was very scarce in passwords created by the 3^{rd} - 5^{th} graders. There was some increasing usage of special characters by the 6^{th} - 8^{th} graders. Many passwords consist of concepts reflecting the current state of the children's lives, e.g., fairy tales, numbers, colors, games, and sports. Few examples from passwords created by the 3^{rd} - 5^{th} graders are: "12345," "Yellow," "PrincessFrog248," and "doggysafesecure." Some passwords created by the 6^{th} - 8^{th} graders are: "Gamehead77," "GameGuy007," "Basketball1130," and "Blue101213." The simplistic nature of passwords is expected since students are progressing on their literacy levels as they age. This is especially true with younger students who are working on mastering their alphabets and numbers. Special characters are such a foreign concept to many young students. This is evidenced by the fact that only 6 special characters appeared in the passwords created by the 3^{rd} - 5^{th} graders, namely, dash (-), period (.), exclamation (!), question (?), at sign (@), and underscore (_), with frequencies all under 1%. The usage of special characters did expand to more types with the 6^{th} - 8^{th} graders: exclamation (!), slash (/), period (.), comma (,), underscore (_), double quote ("), at sign (@), apostrophe ('), left and right parentheses ( ), and caret (^), with frequencies all under 1%.

Despite the awareness shown when discussing the purposes of passwords, the passwords chosen by the children (particularly by the younger age group) were very weak. There were some improvements in the older group, but, as a whole, the passwords were not strong. This suggests that children are not choosing weak passwords because they do not understand the importance of protecting themselves online but because they are either unaware of what constitutes a strong password, or are unable to generate one. There is clearly a need to address how children, particularly in the younger age group, understand and use passwords. When considering the strength of password required it may be worth considering what it is that is being protected and how strong a password is needed. While there have been several investigations into alternative methods of authentication for children, as yet none have been widely adopted. It is important that we do not make passwords "too easy" for children to use leading to their resisting more complicated passwords as adults. The password demands need to challenge children, while still being achievable. Traditional password requirements would suggest that the complexity and strength required should increase as the child's ability develops. However, new password guidelines published by the National Institute of Standards and Technology (NIST) suggests encouraging longer passwords (passphrases) while relaxing complexity requirements (i.e., not requiring a combination of different character types) [34].

VI. CONCLUSION AND FUTURE WORK

Users' password behaviors and experiences across all three stages in the password lifecycle influence users' attitudes and perceptions of password use and requirements. It is important to promote positive user attitudes early on. Users holding positive attitudes towards passwords practice better cyber hygiene, such as creating compliant and strong passwords, writing down passwords less often, suffering less frustration with authentication, better understanding and respecting the significance of security, as compared to users with negative attitudes [14]. This suggests the need to encourage young users to develop positive attitudes and accurate mental models of passwords and authentication.

It is already well known that children are not a homogeneous group. Studies involving children as participants in designing or evaluating new software products will frequently take their literacy or numeracy skills into consideration. This study suggests that, despite growing up with an awareness of cyber security, children are likely to have had widely different experiences and knowledge surrounding passwords. As such, researchers working in the areas of authentication should ensure they measure this existing skill and knowledge base when conducting studies with this population.

It is very important that, when we teach children about passwords, we do not confuse them. There are important differences between security, privacy and safety and it is worth finding ways of communicating these differences to children as they learn about password practice and hygiene. They need to understand the differences between password security, personal information privacy and online safety. We do not want to make children afraid of cyber security and we want them to understand their privacy rights. This is a huge challenge but needs urgent attention.

This paper reports the preliminary findings of a case study from two Midwest schools. We are currently collecting data from other planned US school districts. We plan to perform thorough statistical analysis on the entire dataset of 1500 to 1800 students to further validate the findings in this paper and increase generalizability.

REFERENCES

[1] 3rdknight, "Password complexity for primary school," 2010, edugeek 7 May http://www.edugeek.net/forums/windows/55683-password-complexity-primary-school.html.

[2] A. E. M. Aloufi, "A Cognitive Theory-based Approach for the Evaluation and Enhancement of Internet Security Awareness among Children Aged 3-12 Years," Master's thesis, Rochester Institute of Technology, 2015.

[3] N. Alqahtani, S. Furnell, S. Atkinson, and I. Stengel, "Internet risks for children: Parents' perceptions and attitudes: An investigative study of the Saudi Context," in *Internet Technologies and Applications (ITA)*, Sept 2017, pp. 98–103.

[4] Anonymous, "Biometrics - it's all child's play!" *Biometric Technology Today*, vol. 15, no. 5, pp. 7–8, 2007.

[5] A. L. Bair, "Implementing a digital family policy," March 2015, USA Today.

[6] S. Bauman and H. Pero, "Bullying and cyberbullying among deaf students and their hearing peers: An exploratory study," *Journal of deaf studies and deaf education*, vol. 16, no. 2, pp. 236–253, 2011.

[7] BBC, "Xbox password flaw exposed by five-year-old boy," 2014, http://www.bbc.co.uk/news/technology-268791854April.

[8] I. R. Berson and M. J. Berson, "Children and their digital dossiers: Lessons in privacy rights in the digital age." *International Journal of Social Education*, vol. 21, no. 1, pp. 135–147, 2006.

[9] I. B. Bovina, N. V. Dvoryanchikov, and S. V. Budykin, "Shared meanings about information security of children: An exploratory study," *Procedia-Social and Behavioral Sciences*, vol. 146, pp. 94–98, 2014.

[10] D. Boyd, "How Parents Normalized Teen Password Sharing," 2012, http://www.zephoria.org/thoughts/archives/2012/01/23/ January 23.

[11] C. Chang, "Internet safety survey: Who will protect the children," *Berkeley Tech. Law Journal*, vol. 25, pp. 501–527, 2010.

[12] S. Chaudron, "Young children (0-8) and digital technology," *A qualitative exploratory study across seven countries. Joint Research Centre. European Commission*, 2015.

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

[13] A. D. Cheok, O. N. N. Fernando, and C. L. Fernando, "Petimo: safe social networking robot for children," in *Proceedings of the 8th International Conference on Interaction Design and Children*. ACM, 2009, pp. 274–275.

[14] Y.-Y. Choong, "A cognitive-behavioral framework of user password management lifecycle," in *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 2014, pp. 127–137.

[15] J. Clarke, "The Internet according to kids," *Young Consumers*, vol. 3, no. 2, p. 4552, 2002.

[16] P. E. Coggins III, "Implications of what children know about computer passwords," *Computers in the Schools*, vol. 30, no. 3, pp. 282–293, 2013.

[17] ——, "A pedagogical example of second-order arithmetic sequences applied to the construction of computer passwords by upper elementary grade students," *International Journal of Mathematical Education in Science and Technology*, vol. 46, no. 3, pp. 441–450, 2015.

[18] J. Cole, G. Walsh, and Z. Pease, "Click to enter: Comparing graphical and textual passwords for children," in *Proceedings of the 2017 Conference on Interaction Design and Children*, ser. IDC '17, 2017, pp. 472–477.

[19] ——, "Click to enter: Comparing graphical and textual passwords for children," in *Proceedings of the 2017 Conference on Interaction Design and Children*, ser. IDC '17, New York, NY, USA, 2017, pp. 472–477.

[20] Daily Mail, "Eight in ten children know the passwords and PIN codes to their parents' laptops, phones and computers," May 2013, http://www.dailymail.co.uk/news/article-2328193/.

[21] D. Daramola, "Young Children as Internet Users and Parents Perspectives," Master's thesis, University of Oulu, 2015.

[22] G. Davis, "Bringing up the next generation of digital citizens," 2015, https://blogs.mcafee.com/consumer/family-safety-survey-results/.

[23] T. S. P. D. de Castro, ""Its a complicated situation". Harm in everyday experiences with technology. A qualitative study with school-aged children," Ph.D. dissertation, Universidade do Minho, 2015.

[24] J. F. DeFranco, "Teaching Internet Security, Safety in Our Classrooms," *Techniques: Connecting Education and Careers (J1)*, vol. 86, no. 5, pp. 52–55, 2011.

[25] S. Dredge, "How do I keep my children safe online? What the security experts tell their kids," 2014, the Guardian.

[26] A. Druin, "Lifelong interactions designing online interactions: what kids want and what designers know," *interactions*, vol. 15, no. 3, pp. 42–44, 2008.

[27] G. Fahrnberger, D. Nayak, V. S. Martha, and S. Ramaswamy, "Safechat: A tool to shield children's communication from explicit messages," in *14th International Conference on Innovations for Community Services (I4CS)*. IEEE, 2014, pp. 80–86.

[28] C. Farivar, "This 11-year-old is selling cryptographically secure passwords for $2 each," 2015, http://arstechnica.com/business/2015/10 Ministry of Innovation / Business of Technology.

[29] J. Feng, J. Lazar, L. Kumin, and A. Ozok, "Computer usage by children with down syndrome: Challenges and future research," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 2, no. 3, pp. 13:1–13:44, 2010.

[30] J. Fruth, R. Merkel, and J. Dittmann, "Security warnings for childrens smart phones: A first design approach," in *Communications and Multimedia Security*. Springer, 2011, pp. 241–243.

[31] J. Fruth, C. Schulze, M. Rohde, and J. Dittmann, "E-learning of it security threats: A game prototype for children," in *Communications and Multimedia Security*. Springer, 2013, pp. 162–172.

[32] S. Furnell, "Getting past passwords," *Computer Fraud & Security*, vol. 2013, no. 4, pp. 8–13, 2013.

[33] V. Ginodman, N. Obelets, R. Herkanaidu, R. Reid, and J. Van Niekerk, "Snakes and ladders for digital natives: information security education for the youth," *Information Management & Computer Security*, vol. 22, no. 2, pp. 179–190, 2014.

[34] P. Grassi, E. Newton, R. Periner, A. Regensheid, W. Burr, J. Richer, N. Lefkovitz, J. Danker, Y.-Y. Choong, K. Greene, and M. Theofanos, "Digital identity guidelines: Authentication and lifecycle management," NIST Special Publication, Tech. Rep. 800-63B, 2017.

[35] J. Guan and J. Huck, "Children in the digital age: exploring issues of cybersecurity," in *Proceedings of the 2012 iConference*. ACM, 2012, pp. 506–507.

[36] A. L. Gutnick, M. Robb, L. Takeuchi, and J. Kotler, "Always connected: The new digital media habits of young children," 2011, sesame Workshop and the Joan Ganz Cooney Center.

[37] A. Hajirnis, "Social media networking: Parent guidance required," *The Brown University Child and Adolescent Behavior Letter*, vol. 31, no. 12, pp. 1–7, 2015.

[38] B. A. Handler, "Method and system for child authentication," Mar. 10 2015, uS Patent 8,978,130.

[39] L. Hanna, K. Risden, and K. Alexander, "Guidelines for usability testing with children," *interactions*, vol. 4, no. 5, pp. 9–14, 1997.

[40] B. Holfeld and B. J. Leadbeater, "The nature and frequency of cyber bullying behaviors and victimization experiences in young canadian children," *Canadian Journal of School Psychology*, vol. 30, no. 2, pp. 116–135, 2015.

[41] K. Hundlani, S. Chiasson, and L. Hamid, "No passwords needed: The iterative design of a parent-child authentication mechanism," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '17, New York, NY, USA, 2017, pp. 45:1–45:11.

[42] A. Imran, "A comparision of password authentication between children and adults," Ph.D. dissertation, Carleton University Ottawa, 2015.

[43] L. A. Jackson, R. Samona, J. Moomaw, L. Ramsay, C. Murray, A. Smith, and L. Murray, "What children do on the Internet: Domains visited and their relationship to socio-demographic characteristics and academic performance," *CyberPsychology & Behavior*, vol. 10, no. 2, pp. 182–190, 2006.

[44] C. Johnston, "Social media is parents' greatest online fear, research says," 2014, the Guardian, 12 November.

[45] A. Julka, G. Stehr, D. Parks, and D. Trechter, "Student use of communication technologies – parent/guardian survey report," March 2010, survey Research Center Report 2010/8 http://www.uwrf.edu/SurveyResearchCenter/loader.cfm?csModule=security/getfile&PageID=27599.

[46] L. T. Karklins, "The risks of social networking sites to South Australian high school students," Ph.D. dissertation, Flinders University of South Australia, School of Law., 2013.

[47] J. Kim, K. Kim, J. Park, and T. Shon, "A scalable and privacy-preserving child-care and safety service in a ubiquitous computing environment," *Mathematical and Computer Modelling*, vol. 55, no. 1, pp. 45–57, 2012.

[48] L. Kimmerly and D. Odell, "Children and computer use in the home: Workstations, behaviors and parental attitudes," *Work*, vol. 32, no. 3, pp. 299–310, 2009.

[49] K. Kopeckỳ, "Czech children and facebook–a quantitative survey," *Telematics and Informatics*, vol. 33, no. 4, pp. 950–958, 2016.

[50] ——, "Misuse of web cameras to manipulate children within the so-called webcam trolling," *Telematics and Informatics*, vol. 33, no. 1, pp. 1–7, 2016.

[51] G. Kress, "Before writing: Rethinking the pathway into writing," 1997.

[52] E. Kritzinger, "Enhancing cyber safety awareness among school children in South Africa through gaming," in *Science and Information Conference (SAI), 2015*. IEEE, 2015, pp. 1243–1248.

[53] P. Kumar, S. M. Naik, U. R. Devkar, M. Chetty, T. L. Clegg, and J. Vitak, "'no telling passcodes out because they're private': Understanding children's mental models of privacy and security online," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, pp. 64:1–64:21, Dec. 2017.

[54] D. R. Lamichhane and J. C. Read, "Investigating children's passwords using a game-based survey," in *Proceedings of the 2017 Conference on Interaction Design and Children*, ser. IDC '17, New York, NY, USA, 2017, pp. 617–622.

[55] M. A. Lauri, J. Borg, and L. Farrugia, "Childrens Internet Use and Parents Perceptions of their Children's Online Experience," a study commissioned by the Malta Communications Authority.

[56] P. Y. Lee and Y.-Y. Choong, "Human generated passwords–the impacts of password requirements and presentation styles," in *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 2015, pp. 83–94.

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

[57] W. Loban, *The language of elementary school children.* National Council of Teachers of English, Champaign, IL, 1963.

[58] B. Lorenz, K. Kikkas, and A. Klooster, ""The Four Most-Used Passwords Are Love, Sex, Secret, and God": Password Security and Training in Different User Groups," in *Human Aspects of Information Security, Privacy, and Trust*. Springer, 2013, pp. 276–283.

[59] S. MacFarlane, J. Read, J. Höysniemi, and P. Markopoulos, "Half-day tutorial: Evaluating interactive products for and with children," in *Interact*, 2003, pp. 1027–1028.

[60] T. Mendori, M. Kubouchi, M. Okada, and A. Shimizu, "Password input interface suitable for primary school children," in *Computers in Education, 2002. Proceedings. International Conference on*. IEEE, 2002, pp. 765–766.

[61] D. J. Meter and S. Bauman, "When sharing is a bad idea: the effects of online social network engagement and sharing passwords with friends on cyberbullying involvement," *Cyberpsychology, Behavior, and Social Networking*, vol. 18, no. 8, pp. 437–442, 2015.

[62] F. Mishna, M. Saini, and S. Solomon, "Ongoing and online: Children and youth's perceptions of cyber bullying," *Children and Youth Services Review*, vol. 31, no. 12, pp. 1222–1228, 2009.

[63] A. Mubarak, O. Judy, and Z. Harrison, "Bridging the gap between adolescents and adults: A case study on experiences of teenagers while using social networking sites and their willingness to seek adult help," *International Journal of Computer Applications*, vol. 102, no. 6, pp. 36–42, 2014.

[64] B. Nansen, K. Chakraborty, L. Gibbs, C. MacDougall, and F. Vetere, "Children and Digital Wellbeing in Australia: Online regulation, conduct and competence," *Journal of Children and Media*, vol. 6, no. 2, pp. 237–254, 2012.

[65] L. Pangrazio and N. Selwyn, "'My data, my bad'-Young peoples personal data understandings and (counter) practices," in *Proceedings of the 8th International Conference on Social Media & Society*, July 2830, Toronto, ON, Canada, 2017.

[66] P. Paul, "Cracking teenagers online codes," *The New York Times*, vol. 20, 2012.

[67] Y. Poullet, "e-Youth before its judges–Legal protection of minors in cyberspace," *Computer Law & Security Review*, vol. 27, no. 1, pp. 6–20, 2011.

[68] J. C. Read and R. Beale, "Under my pillow: designing security for children's special things," in *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*. British Computer Society, 2009, pp. 288–292.

[69] J. C. Read and B. Cassidy, "Designing textual password systems for children," in *Proceedings of the 11th International Conference on Interaction Design and Children*. ACM, 2012, pp. 200–203.

[70] R. Reid and J. van Niekerk, "A cyber security culture fostering campaign through the lens of active audience theory," in *Proceedings of the Ninth International Symposium on Human Aspects of Information Security & Assurance*, Lesvos, Greece, 1-3 July 2015, pp. 34–44.

[71] K. Renaud, "Web authentication using Mikon images," in *Privacy, Security, Trust and the Management of e-Business, 2009. Congress'09. World Congress on*. IEEE, 2009, pp. 79–88.

[72] K. Renaud and J. Maguire, "Regulating access to adult content (with privacy preservation)," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 4019–4028.

[73] M. Ribble, *Raising a Digital Child: A Digital Citizenship Handbook for Parents*, 1st ed. International Society for Technology in Education, 2009.

[74] M. Richtel, "Young, in love and sharing everything, including a password," 17 January 2012, http://www.nytimes.com/2012/01/18/us/teenagers-sharing-passwords-as-show-of-affection.html.

[75] K. Rim, "Investigation and analysis of password use for the betterment of privacy protection education in middle and high school students," *Advanced Science and Technology Letters*, vol. 103, pp. 72–76, 2015.

[76] K. Rim and S. Choi, "Analysis of password generation types in teenagers–focusing on the students of jeollanam-do," *International Journal of u-and e-Service, Science and Technology*, vol. 8, no. 9, pp. 371–380, 2015.

[77] J. A. Rode, "Digital parenting: designing children's safety," in *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*. British Computer Society, 2009, pp. 244–251.

[78] A. Samuel, "Your Weakest Security Link? Your Children," http://www.wsj.com/articles/your-weakest-security-link-your-children-1429499471 Wall Street Journal, 19 April.

[79] M. Sharples, R. Graber, C. Harrison, and K. Logan, "E-safety and web 2.0 for children aged 11–16," *Journal of Computer Assisted Learning*, vol. 25, no. 1, pp. 70–84, 2009.

[80] C. A. Shoniregun and A. Anderson, "Is Child Internet Access a Questionable Risk?" *Ubiquity*, vol. 2003, no. September, pp. 4:1–4:1, Sep. 2003.

[81] E. Stobert and R. Biddle, "Authentication in the home," in *Workshop on Home Usable Privacy and Security*, 2013.

[82] J. D. Sutter, "Survey: 70% of teens hide online behavior from parents," 2012, http://www.cnn.com/2012/06/25/tech/web/mcafee-teen-online-survey/.

[83] T. Swist, P. Collin, J. McCormack, A. Third, and W. Australia, *Social media and the wellbeing of children and young people: A literature review*. Commissioner for Children and Young People, 2015.

[84] İ. Tanrıkulu, S. A. Altun, Ö. E. Baker, and O. Y. Güneri, "Misuse of icts among turkish children and youth: A study on newspaper reports," *International Journal of Human Sciences*, vol. 12, no. 1, pp. 1230–1243, 2015.

[85] M. Teimouri, M. S. Hassan, J. Bolong, A. Daud, S. Yussuf, and N. A. Adzharuddin, "What is upsetting our children online?" *Procedia-Social and Behavioral Sciences*, vol. 155, pp. 411–416, 2014.

[86] A. D. Thierer, "Kids, privacy, free speech & the Internet: Finding the right balance," *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1909261.*, 2011.

[87] C. A. Tomlinson, *How to differentiate instruction in mixed-ability classrooms*, 2nd ed. Association for Supervision and Curriculum Development, Alexandria, Virginia., 2001.

[88] A. Trinneer, "Teaching children about Internet safety: An evaluation of the effectiveness of an interactive computer game," Master's thesis, School of Psychology, University of Ottawa (Canada), 2006.

[89] S. J. Tsim, "Internet safety education: Information retention among middle school aged children," Master's thesis, San José State University, 2006.

[90] A. Uhl and P. Wild, "Comparing verification performance of kids and adults for fingerprint, palmprint, hand-geometry and digitprint biometrics," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems,(BTAS)*, 2009, pp. 1–6.

[91] S. Vandoninck and L. d'Haenens, "Children's online coping strategies: Rethinking coping typologies in a risk-specific approach," *Journal of adolescence*, vol. 45, pp. 225–236, 2015.

[92] T. Velki, K. Solic, V. Gorjanac, and K. Nenadic, "Empirical study on the risky behavior and security awareness among secondary school pupils - validation and preliminary results," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2017, pp. 1280–1284.

[93] M. Villano, "Text unto others... as you would have them text unto you," September 2008, https://thejournal.com/articles/2008/09/01/text-unto-others-as-you-would-have-them-text-unto-you.aspx.

[94] J. Vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven, "Reconstructing the giant: On the importance of rigour in documenting the literature search process." in *ECIS*, vol. 9, 2009, pp. 2206–2217.

[95] J. Weed, "Online monitoring: Do you know your child's passwords?" 2011, 28 March https://www.parentmap.com/article/online-monitoring-do-you-know-your-childs-passwords.

[96] H. Zaikina-Montgomery, "The dilemma of minors' access to adult content on the Internet: A proposed warnings solution," Ph.D. dissertation, University of Nevada, Las Vegas, 2011.

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

APPENDIX: Survey Alignment Matrix

Objective	Category	Category Definition	Survey Items
RQ1: Assess children's use of computers	Usage	Extent to which children use computers	Where do you use computers? • What types of computers do you use? • Where do you use computers? • About how much time do you spend on computers each day during the week? • About how much time do you spend on computers during the weekend? • What do you do when you go on the computer? (list of activities)
RQ1: Assess children's use of computers	Usage	Extent to which children use passwords	• How many passwords do you have? • I use passwords to login into (list of activities). • How many times a day do you use or enter your passwords?
RQ2: Assess children's knowledge of passwords	Knowledge	Extent to which children understand purpose and appropriate security practices with passwords	• Where did you learn about good password use? • Let's talk about your passwords: ∗ Do you share your password with friends? ∗ Do you use the same password for everything? ∗ Do you write your password down on paper? ∗ Do you change your passwords? ∗ When do you change your passwords? ∗ When you finish with the computer do you log out? • List up to 3 reasons why we need passwords?
RQ2: Assess children's knowledge of passwords	Perceptions	Extent to which students are comfortable with passwords and password practices	• What do you think of passwords (scales of agreement) ∗ It is easy to create my password ∗ It is easy to create many different passwords ∗ It is easy to remember my passwords ∗ It is easy to type in my passwords ∗ I wish there was another way to login besides passwords ∗ I have too many passwords
RQ3: Assess children's current password behaviors	Memory	Extent to which children are self-reliant with respect to passwords	• How do you pick your password? • How do you remember your passwords? • Do you help your family members with passwords? (6^{th} - 12^{th} graders only) • Let's say you just got a new game to play on the computer but you need a password to use it. Please make up a new password for that game.
• Gender, age and school differences.	Demographics	Age, gender, which school which grade	Are you a girl or boy, prefer not to answer? • How old are you? • What grade are you in? • Where do you go to school? • Which city do you live in?

Choong, Yee-Yin; Theofanos, Mary; Renaud, Karen; Prior, Suzanne.
"Case Study - Exploring Children's Password Knowledge and Practices."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

RFC: DLMF Content Dictionaries

Bruce R. Miller
National Institute of Standards and Technology
Gaithersburg, MD, USA
bruce.miller@nist.gov

1 Overview

The Digital Library of Mathematical Functions (DLMF[1]) covers the definitions and properties of a wide variety of special functions with applications in physics and engineering — several hundred, depending on how you count them. Although initially focused as a resource for human readers, the DLMF's long-term goal is to support machine-readable content to enable interoperability between other digital libraries, computer algebra and theorem proving systems.

But, the long history of special functions across and among different communities of interest has led to the potential for confusion and error. When different practitioners speak about apparently the same function, their different histories and conventions may include different assumptions of scalings, arguments (order or definition), branch cuts, assumed values in special cases, and so on. A simple example is the Jacobian elliptic function sn seen as a function either of the modulus k or the parameter $m = k^2$; each form is preferred for certain purposes. While neither party is necessarily *wrong*, failure to account for their differences in communications is guaranteed to lead to error.

It is thus critical for interoperability between systems to establish each system's conventions and assumptions. This is true enough for a human reader transcribing DLMF's results into a computer algebra system, but all the more so when these processes are automated and hidden from view. Ideally these differences can be formalized to the extent that enables automatic conversion of formula across the different views. Indeed, the World Digital Mathematics Library[2] has founded an effort to develop such a Special Function Concordance.

Towards these ends, this note presents a proposed set of (virtual) OpenMath[3] Content Dictionaries (CD) to characterize the choices made in the DLMF.

2 Organization and Naming Conventions

An unexpected challenge was an organization and naming of the CDs and symbols in a fashion appropriate to OpenMath applications. The functions can be grouped according to mathematical or historical features, or applications. The proper names of functions can become quite verbose with strings of significant adjectives before they become sufficiently unique. Consider "Legendre's incomplete elliptic integral of the first kind" and then add modifiers such as "zeros of the derivatives of". Given that some functions are ubiquitous while others are truly esoteric, one would even hope for a Huffman-type encoding.

Yet, the functions have already been grouped into chapters in a way appropriate for the DLMF's purposes. Moreover, each function has a unique LaTeX macro defined for it to simplify the markup and preserve the semantics during conversion to web formats [4]. For example the two functions mentioned above have, macros `\Jacobiellsnk` (encoding "the Jacobian elliptic function sn, of modulus k") and `\incellintFk` (encoding "(Legendre's) incomplete elliptic integral (of the first kind) of modulus k"; See Appendix A for details). While these

In: O. Hasan, J. Davenport, M. Kohlhase (eds.): Proceedings of the 29th OpenMath Workshop, Hagenberg, Austria, 13-Aug-2018, published at http://ceur-ws.org

[1] `https://dlmf.nist.gov/`
[2] `https://www.mathunion.org/ceic/library/world-digital-mathematics-library-wdml`
[3] `https://openmath.org/`
[4] a DLMF Macro set, to be released, is under development

Miller, Bruce.
"RFC: DLMF Content Dictionaries."
Paper presented at 11th Conference on Intelligent Computer Mathematics CICM 2018, Hagenberg, Austria. August 13, 2018 - August 17, 2018.

Table 1: Types used in the special function type signature, where $\mathcal{T}$ stands for any arbitrary set or type.

Notation	*Meaning*
$\mathcal{T} \to \mathcal{T}'$	function (mapping) from type $\mathcal{T}$ to $\mathcal{T}'$
$\mathcal{T} \times \mathcal{T}'$	product set of multiple types
$\mathbb{R}$	the set of real numbers (excluding ∞)
$\mathbb{C}$	the set of complex numbers (excluding ∞)
$\mathbb{Z}$	the set of integers
$\mathbb{Z}^+$	the set of integers > 0
$\mathbb{Z}^*$	the set of integers ≥ 0
$\mathbb{Q}$	set of rational numbers
$\mathbb{D}$	the complex numbers in the open unit disc, $\lvert z\rvert < 1$
$\{a_0, \ldots\}$	one of a finite set (of symbols)
$\mathcal{T}^n$	n-tuples with elements of type $\mathcal{T}$ (e.g. $\mathbb{R}^2$ for pairs of reals)
$\mathcal{T}^\bullet$	tuples with elements of type $\mathcal{T}$, unspecified length
$\mathcal{T}^n$	vectors of dimension n, with elements of type $\mathcal{T}$
$\mathcal{T}^\bullet$	vectors with elements of type $\mathcal{T}$, unspecified dimension
$\mathcal{T}^{n\times m}$	$n \times m$ matrices with elements of type $\mathcal{T}$
$\mathcal{T}^{\bullet\times\bullet}$	matrices with elements of type $\mathcal{T}$ (unspecified dimension)
$\mathbf{L}$	lattices in the complex plane (in the sense of elliptic functions)

may not roll off the tongue, they are unique, reasonably type-able and blend with TEX's macro conventions. And while this organization and naming may not be optimized for OpenMath purposes, it seems better to reuse the one scheme than to introduce redundant ones.

We have therefore followed DLMF's organization for the primary grouping of functions. The most important functions in a chapter are covered in a base CD, such as `DLMF_BS` (for Bessel functions). In most cases, progressively esoteric functions are grouped into subcategories according to: generalizations (`DLMF_BS_gen`), q-analogs (`DLMF_BS_q`), magnitudes, zeros, matrix argument and so on, as well as some special case such as `DLMF_GH_Appell` for Appell functions.

3 Characterizing the Functions

The more fundamental challenge is to properly characterize the functions. This, of course, is exactly what any proper 'definition' ought to be. But here the point is that the definition be sufficiently complete, explicit and formalized, to enable easily determining the equivalancy of functions from different systems. Ultimately, the goal would be to enable automatic conversion between, for example, the two different 'flavors' of elliptic functions, sn.

At this stage of development, we are providing URLs as the definitions of each function, being pointers into the DLMF where the definition is to be found. This is obviously an informal definition, and may require digging for some details. Definitions may be either explicit or implicit (such as a function defined by a differential equation along with boundary conditions).

Additionally, we have provided a simple type signature for each function to characterize its domain and range (See Table 1). Note that in many cases functions are undefined for isolated values of some arguments, e.g. singularities; these cases are not always reflected in the current signatures. Other properties, such as branch cuts, multivaluedness, have not yet been made explicit.

4 Conclusions and Request for Comments

We have provided here a catalog of the special functions covered by the DLMF — a set of informal, virtual Content Dictionaries. It should serve as a reasonable starting point for establishing a concordance between the sets of functions covered by the several interested parties. This is a continuing process; our CDs will continue to be refined and gradually extended and formalized as needed.

The current status can be found at `https://math.nist.gov/~BMiller/DLMF-CDS/`, where also a JSON encoding of the data may be downloaded for processing.

We welcome suggestions about which features and characteristics function are important to the notion of a concordance, as well as how best to encode and formalize that information. Any other comments about the catalog are also welcome.

Acknowledgements: The author would like to thank Patrick Ion, Howard Cohl and Florian Rabe for constructive comments.

A DLMF Macro Naming conventions

Briefly, the names of the various mathematical function macros are derived from the descriptive 'Proper Name' of the function according to:

$$macro \equiv \backslash\ prefix^{*}\ name\ class^{?}\ symbol^{?}\ suffix^{*}$$

The *name* is the 'conventional' name or based on the "inventor's" name. The *class* indicates function (generally omitted), integrals, polynomials, and so on. The *symbol* is the latinized form of the notation, upper or lower case as appropriate. The *prefix* modifier includes *all* significant characteristics that may distinguish functions (e.g. 'modified Bessel' vs. simply 'Bessel'). The *suffix* generally indicates limitations or special cases regarding arguments. The abbreviations used for *prefix*, *class* and *suffix* are given in Table 2. For predictability, we avoid abbreviating people's names.

Table 2: Abbreviations used for DLMF macros

class	*Meaning*
	function (omitted by default)
char	characteristic
eigval	eigenvalues
eigvec	eigenvectors
int	integral
mod	modulus
number	number
phase	phase (or phase shift)
poly	polynomial
sum	sum
sym	symbol
trans	transform
wave	wavefunction

suffix	*Meaning*
imag	imaginary argument or order
k	elliptic functions of k, modulus
m	elliptic functions of parameter $m = k^2$
mat	matrix argument
real	of real argument or order
invar	on invariants (Weierstrass)
latt	on lattice (Weierstrass)
q	functions of q, nome
tau	functions of τ

prefix	*Meaning*
a	arc, inverse (circular functions)
A	arc, multi-valued-inverse
aff	affine
ass	associated
aux	auxiliary
big	big
canon	canonical
comp	complete
ccomp	complete complementary
cont	continuous
cusp	cuspoid
deriv	derivative(s) of
diff	differential
diffr	diffraction
dil	dilated
disc	discrete
div	dividing
dual	dual
ell	elliptic
env	envelope of
exp	exponential
gen	general | generalized
hyper	hyperbolic | hypergeometric
inc	incomplete
inv	inverse
irreg	irregular
little	little
log	logarithm(ic)
mod	modified (or modular?)
multivar	multivariate
n	number of
norm	normalized or normalization
para	parabolic
per	periodic
q	q-variant of
rad	radial
reg	regular
rest	restricted
sc	scaled
shift	shifted
sph	spherical | spheroidal
sym	symmetric
umb	umbilic
usph	ultraspherical
z	zeros (of)

MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation

Haiying Guan[1], Mark Kozak[2], Eric Robertson[2], Yooyoung Lee[1], Amy N. Yates[1], Andrew Delgado[1], Daniel Zhou[1], Timothee Kheyrkhah[1], Jeff Smith[3], Jonathan Fiscus[1]

[1]National Institute of Standards and Technology, [2]PAR Government, [3]University of Colorado Denver

haiying.guan@nist.gov, {mark_kozak, eric_robertson}@partech.com, {yooyoung.lee, amy.yates, andrew.delgado, daniel.zhou, timothee.kheyrkhah}@nist.gov, jeff.smith@ucdenver.edu, jonathan.fiscus@nist.gov

Abstract

We provide a benchmark for digital Media Forensics Challenge (MFC) evaluations. Our comprehensive data comprises over 176,000 high provenance (HP) images and 11,000 HP videos; more than 100,000 manipulated images and 4,000 manipulated videos; 35 million internet images and 300,000 video clips. We have designed and generated a series of development, evaluation, and challenge datasets, and used them to assess the progress and thoroughly analyze the performance of diverse systems on a variety of media forensics tasks in the past two years.

In this paper, we first introduce the objectives, challenges, and approaches to building media forensics evaluation datasets. We then discuss our approaches to forensic dataset collection, annotation, and manipulation, and present the design and infrastructure to effectively and efficiently build the evaluation datasets to support various evaluation tasks. Given a specified query, we build an infrastructure that selects the customized evaluation subsets for the targeted analysis report. Finally, we demonstrate the evaluation results in the past evaluations.

1. Introduction

The explosion of media storage, transmission, editing, and sharing tools has the potential to foster an increase in tampered data and challenge the traditional trust in visual media. The creation, modification, and distribution of digital media are extremely simple to use for even inexperienced users and require only minimal effort. In addition, with developments in the advent of new techniques such as Generative Adversarial Networks (GANs, "deepfakes") [1] and Computer Generated Imagery (CGI) [2], it becomes increasingly challenging for existing media forensic technologies[3]-[10]to identify the integrity, trustworthiness, and authenticity of visual content.

In order to facilitate the development of media forensics research, we started work on the Media Forensics Challenge (MFC) Evaluation for the DARPA MediFor program [11] in 2015. We design, collect, annotate, and assemble a series of comprehensive digital forensic databases for the evaluation of media forensic technologies. The benchmark data contains four major parts: (1) 35 million images and 300,000 video clips downloaded from the internet with their characteristics and labels; (2) up to 176,000 pristine, self-collected, high provenance (HP) images and 11,000 HP videos; (3) approximately 100,000 manipulated images covering over 100 manipulation types produced by professional manipulators from approximately 5,000 image manipulation journals with manipulation history graphs and annotation details; 4,000 manipulated videos and over 500 video manipulation journals; (4) a series of evaluation datasets with reference ground-truth to support several challenge tasks in media forensics challenge evaluations from the last two years.

In this paper, we first survey existing datasets and discuss the special characteristics and challenges of media forensic data collection. Then based on the diversity and complexity of media forensic applications, we propose the following tasks [12][13]: Manipulation Detection and Localization: to detect if an image/video has been manipulated (MD), and if so, where it is manipulated (MDL); Splice Detection and Localization: to detect if a region of a given potential donor image has been spliced into a probe image (SD); if so, where are the regions in both images (SDL); Provenance Filtering (PF) and Provenance Graph Building (PGB): to reconstruct an image's phylogeny graph given a 'world' image pool; Camera Verification (CV): to verify if an image/video probe is from a claimed camera sensor; and Event Verification (EV): to verify if an image is from a claimed event. Detection systems are measured by the Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC), localization systems are measured by the Matthews Correlation Coefficient (MCC), PF systems are measured by Recall, and PGB systems are measured by the Node Link Overlap Similarity Metric (SimNLO). Afterwards, we present the approach to building evaluation datasets for different evaluation tasks and demonstrate the results from the last two years' evaluations.

The major contributions are: (1) we propose and demonstrate a methodology for MFC evaluation data design; (2) we build large-scale media forensics evaluation benchmark datasets for quantitative media forensics system performance evaluations. The evaluation data can be directly used for existing task evaluations. The world data, HP data, and manipulation data can be easily customized to other evaluation tasks; (3) we provide a description about the data, metadata, training data, and ground-truth evaluation reference data of our datasets, which includes but is not limited to the capture camera/device's model and identity, the step-by-step manipulation operations with parameters, the manipulation history graph (i.e. journal), the localized manipulated region for every step, the semantic category, the model and parameters of the special filters (e.g. GAN, jpeg anti-forensic filter, social media laundering etc.); (4) we present the state-of-the-art media forensics system evaluation results and their progress over the past two years.

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

2. Related Work[1]

Many general benchmark media datasets are available in literature: UCID (2003) [14] and ImageCLEF (2004-2013) [15] for image retrieval; Caltech-256 (2007) [16], 80 million tiny images (2008) [17], Mammal Image (2008) [18], PASCAL (2008-2015) [19], ImageNet (2009) [20], SUN (2010) [21], Kitti (2013) [22], and Microsoft COCO (2014) [23] for object and scene detection, segmentation, and recognition; TRECVid (2000-2006) [24], Event recognition video dataset (2011) [25], YFCC100M (2015-2016) [26], MediaEval (2013-2017) [27] for multimedia research. Unfortunately, none of them could be directly used for media forensics evaluation purposes due to lack of manipulated media or sufficient annotation. Given a media from any of the above datasets, we would not know if it was manipulated. Furthermore, it is impractical to do post-annotation for manipulation metadata.

Recently, several datasets were collected specifically for media forensics research. The EU REWIND (REVerse engineering of audio-VIsual coNtent Data) (2011-2013) [28] digital forensics project focuses on digital watermarking, passive image authentication, and capture source identification (camera PRNU etc.). The 'Realistic' dataset contains 69 manually manipulated fake images and 69 original images. The 'Synthetic' dataset contains 4800 automatically manipulated images. 200 images from a Nikon D60 camera have been released. The dataset is small and only small portions of it are available to public.

The First Image Forensics Challenge (2013) [29], an international competition organized by the IEEE Information Forensics and Security Technical Committee (IFS-TC), collected thousands of images of various scenes, both indoors and outdoors with 25 digital cameras. We use this dataset as our first reference, and expand the design's breadth and depth in several aspects: scale, manipulation types and graph, annotations, PRNU training data etc.

Some forensic databases target particular manipulation operations. The Columbia automatically-spliced image database (2004) [30] has two parts: a grayscale image dataset with 933 authentic and 912 spliced 128×128 pixel-grayscale image blocks, extracted from images in CalPhotos [31], and a color image dataset with 183 authentic uncompressed color block images and 180 spliced uncompressed color block images. CASIA's Image Tampering Detection Evaluation Database (2013) [32] focuses on splicing and tampering. CASIA v1.0 has 800 authentic and 921 spliced 384×256 images. CASIA v2.0 contains 7,491 authentic and 5,123 tampered images. The CoMoFoD dataset (2013) [33] designed for copy-move forgery detection consists of 260 forged image sets. Each set includes a forged image, two masks and its original image. The manipulations include translation, rotation, scaling, distortion, and postprocessing such as JPEG compression, blurring, noise adding, color reduction etc. The MICC F220, MICC F2000 (2011) [34] and FAU-Erlangen Image Manipulation Datasets (2012) [35] are other copy-move datasets. The Rebroadcast dataset (2018) [36] contains 14,500 large diverse rebroadcast images captured by screen-grabs from 234 displays, scanning printed photos using 173 scanners, or re-photographing displayed or printed photos with 282 printers and 180 recapture cameras. UMDFaces (2016) [37] has 367,888 annotated faces of 8,277 subjects. About 115,000 images have their key point annotations verified by humans. The UMD face swap dataset contains tampered faces created by swapping one face with another using multiple face swapping apps. The VISION dataset (2017) [38] contains 11,732 native images, which are then shared through Facebook and WhatsApp resulting in a total of 34,427 images, and 648 native videos, which are shared through YouTube and WhatsApp, resulting in a total of 1,914 videos. The FaceForensics dataset (2018) [39] has about a half million edited images (from over 1000 videos at various quality levels) using a state-of-the-art face editing approach, and annotated with classification and segmentation references. Those datasets are limited to only single or several manipulation types like splicing, social app, copy-move, rebroadcast, or face manipulations etc.

Other datasets are collected for other purposes. Break Our Steganographic System (BOSS) (2011) [40] is for steganalysis challenge evaluation. Two datasets are created: BOSSBase and BOSSRank, BOSSBase was composed of 9,074 never-compressed cover images coming from 7 different cameras, and created from full-resolution color images in RAW format. The BOSSRank has 1,000 512×512 grayscale images. Several datasets are designed for source identification, that is, to verify the trust and authenticity of data and the devices that create it. Purdue Sensor and Printer Forensics (PSAPF) Dataset (2008) [41] provides an overview of current characterization techniques for 5 scanners and 21 printers. Goljan et al. (2009) [42] presented one million images collected from Flickr, spanning 6,896 individual cameras covering 150 commerce models. The Dresden image database (2010) [43] contains 14,000 high resolution images from 73 digital cameras covering 25 camera models. The images are collected from different scenes with two additional sets of auxiliary images for special use in camera Photo Response Non-Uniformity (PRNU) studies. RAISE (RAw ImageS datasEt) (2015) [44] is a collection of 8,156 high-resolution raw images using three Nikon devices. The images are taken at very high resolution and saved in an uncompressed format. The images cover a wide variety of semantic content, subjects, scenarios, and technical parameters, and are properly

[1] Any mention of commercial products or reference to commercial organizations in this report is for information only; it does not imply recommendation or endorsement by NIST nor does it imply that the products mentioned are necessarily the best available for the purpose.

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

annotated with 7 category labels. These datasets are valuable for our Camera Verification task only.

Despite the availability of the existing datasets, there is a need for a sufficiently large, representative benchmark dataset containing a wide range of realistic media manipulations with detailed phylogeny graphs and ground-truth references for media forensic evaluation to push the state-of-the-art forward.

3. Media Forensic Dataset Design

Our evaluation, besides answering the general question: "How well do the state-of-the-art forensic systems perform?", aims to answer deeper questions such as: (1) What major factors affect the performances? (2) Accounting for the diversity and specialization of media forensics technologies, which systems are suitable for which situations? To answer such questions, it is insufficient merely to have ground-truth about whether the media is manipulated. The data, metadata, manipulation data, and reference data are crucial for the evaluation. The row headers of Table 1 denote data/metadata availability; the row index labels show the tasks and evaluation report availability given the data/metadata references. It shows that more tasks could be evaluated if more reference resources are available. The quality and quantity of the analysis reports depends on the availability of metadata. With well-structured data, metadata, and reference data, we can provide detailed reports and comparisons using factor analysis, which can help us better understand system performance, promote system development, and provide directions for future development and evaluation.

Table 1: Evaluation capability vs. data availability.

support more tasks and more detailed evaluation report	Report contents	Task	Manipulated ?	Probe Ref. Mask	Donor Ref. Mask	Meta data	PRNU Train Data	Manip History Graph
	Det. Scores (ROC, AUC) on full dataset	MD	✓	×	×	×	×	×
	Det. Scores; Localization score (MCC) on full dataset	MDL	✓	✓	×	×	×	×
	For both probe and donor: Det. Scores; Loc. score on full dataset only	MDL SDL	✓	✓	✓	×	×	×
	For both probe & donor, Det. & Loc. score; Factor analysis on both full dataset and subsets by selective scoring using defined metadata queries	MDL SDL	✓	✓	✓	✓	×	×
	For probe & donor, Det. & Loc. score; Factor analysis on both full dataset and subsets by selective scoring using defined metadata queries	MDL SDL CID	✓	✓	✓	✓	✓	×
	For probe & donor, Det. & Loc. score; Factor analysis on both full dataset and subsets by selective scoring using defined metadata queries; PF Recall; PGB SimNLO;	MDL SDL CID PF, PGB	✓	✓	✓	✓	✓	✓

3.1. Data Collection Challenges

Besides the general challenges of computer vision dataset collection, the major challenges for media forensics datasets are: (1) Highly diverse research topics involving multidisciplinary areas: multimedia security, computer forensics, image processing, computer vision, imaging, and signal processing. Technologies include but are not limited to JPEG artifact detection, crop/contrast/clone/splice detection, lighting/shadow/reflection consistency, physical/semantic consistency, Electrical Network Frequency (ENF), PRNU, and audio/video person ID consistency. The evaluation of different types of media forensics systems requires different types of meta, training, testing, and reference data. In addition, different systems may only work with particular constraints. For example, face systems work only on a face region, clone detectors only work well on cloning, which increases the complexity of the evaluation metadata and infrastructure. (2) Intrinsically high dimensionality: besides image/video dimensions and their metadata (EXIF, camera ID etc.), we noticed that the manipulation operations and history of a media are very important for media forensics research. The history generates additional factors which affect system performance. In addition, the manipulation was often done with a purpose and the semantic meaning of the manipulation presents yet more factors. (3) High complexity and cost for structured data and metadata collection, manipulation, and annotation. (4) "Curse of dimensionality" issues: to better understand the system performance, and to do factor analysis (apple-to-apple) comparisons, we need systematically structured orthogonal fractional factorial data. The dataset size increases exponentially as factors increase. (5) The post-annotation approach, which is used for biometric or video analytics evaluation ground-truth data generation, does not work well in the media forensics domain. Given a media, even forensics experts find it difficult to deduce whether it is manipulated, let alone compose a step-by-step description of the manipulation operations and their corresponding manipulated regions used for evaluation.

3.2. Dataset Design and Solutions

Given the challenges listed above, we propose the following approaches and solutions:

(i) A set of sufficiently large and publicly available datasets are collected, annotated, and manipulated. We hired professional manipulators using various media editing software and tools to produce manipulated media suited to real-world applications.

(ii) We proposed a manipulation history graph representation to capture the structured manipulation data, operations, metadata, reference mask etc.

(iii) We designed and developed the manipulation Journaling Tool (JT) [45], a software application that assists us in generating and annotating the data, metadata, reference data, and reference ground-truth mask collection effectively and efficiently. The JT integrates the functions such as semantic and metadata collection, automatic and human annotations, manipulation reference mask generation, manipulation operation data collection, and automatic/semi-automatic manipulation etc. The JT has both online and offline functions to support data collection, generation, annotation and verification based on the collection requirements.

(iv) We have developed an automatic journaling tool (AutoJT), which generates manipulated media and its accompanying journal from non-manipulated media.

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

(v) We have also developed an extended journaling tool (ExtendedJT), which extends and journals partial or complete manipulated media from existing journals, for factor analysis. It automatically generates a large number of manipulated images/videos given a manipulation operation filter and their parameters.
(vi) We designed the data collection and evaluation infrastructure which shares the data among different evaluation tasks to reduce the data collection cost.
(vii) Because different systems may work on different test sets, in addition to testing systems on all testing data, we also designed and developed a selective scoring evaluation infrastructure to dynamically extract a specified subset from the whole test set for selective evaluation.

4. MFC Data Collection

4.1. Images and Videos

World Data: Publicly available imagery acquired off the internet is referred to as World Data. To date the corpus contains 35 million images and 300,000 videos of World Data. It is anticipated that the program will have downloaded 45 million images and 450,000 video clips from the internet. The initial collection of World Data was random, to be used as clutter in evaluations. After the initial collection, we focused on maximizing the diversity of the camera models. Over 500 distinct camera models are represented.

HP Data: At the time this paper was written a total of 176,000 High Provenance (HP) images and 11,000 HP videos were collected by our team. An additional 35,000 photographs and 3,500 videos have been planned to be collected. HP data is always collected on physical devices which team members have physical access to, where all relevant device-intrinsic parameters are known and recorded. Using HP data ensures that all manipulations occur on images with no previous manipulations and avoids copyright infringement. All HP data collected is released under the Creative Commons 0 (CC0) license. CC0 effectively releases the images into the public domain.

PRNU Training Data: Several hundred cameras have been enrolled. These largely consist of moderate to high end Digital Single Lens Reflex (DSLR) and mirrorless cameras. To date the camera database has 574 distinct HP cameras enrolled in the database. Enrolling an HP camera in the database is a multistep process. It begins with the collection of Photo Response Non-Uniformity (PRNU) training data. Image PRNU data is collected with a perfectly diffuse practical light source to evenly illuminate the sensor. Ideally images with pixel saturation at 80% and 50%, and then with a lens cap on are collected at each resolution the camera is capable of. After PRNU data collection the user completes a camera enrollment form to record the owner of the camera, an inventory control ID, make, model, and several other metadata fields. Mobile devices with front and rear facing cameras are further identified by their primary camera (non-selfie) and the secondary (selfie) camera. Most cameras have several hundred, sometimes reaching over 2,000 images per camera.

To facilitate the discovery and management of the media, the "MediFor Browser", a PostgreSQL database and web-based interface, has been developed.

4.2. Manipulation Data and Annotations

Manipulated media is accompanied by a journal of the sequential manipulations executed to produce the media. We describe each manipulation with a software agnostic description called an operation, generalizing the behavior of the manipulation. Along with the operation name, each manipulation description is accompanied by the software name and version used to execute the manipulation, generalized parameters providing details of the operation, and semantic annotations describing the purpose of the manipulation in the context of a group of manipulations.

4.3. Journaling Tool (JT) for Manipulation Recording

The JT [45] collects and administers a sequential record of manipulations applied to non-manipulated media to produce one or more complete manipulated products (Figure 1 (a)). Given the detailed record, the JT provides a set of operational checks for accuracy and unintended side effects. The checks ensure each manipulation is discrete. For example, a common mistake in cropping images is to realign pixels, thus polluting the crop operation with interpolation and resizing.

The journal includes non-manipulated base media, manipulated final media, media captured after each discrete manipulation and the data capturing relevant changes in the media for each manipulation. The last kind of data includes metadata and per-pixel indicators in the form of manipulation masks. Metadata is often a relevant indication of manipulation, such as GPS coordinates, time of day and camera-model adjustments. Retaining all relevant media serves journal extension, in which each step of the manipulation sequence becomes a launching point for another set of manipulations.

5. MFC Evaluation Tasks

Due to the nature and diversity of media forensics, one evaluation task cannot cover all applications. We designed 5 evaluation tasks (MDL image and video, SDL, EV, PF, and PGB), and 2 challenges (CV and GAN).

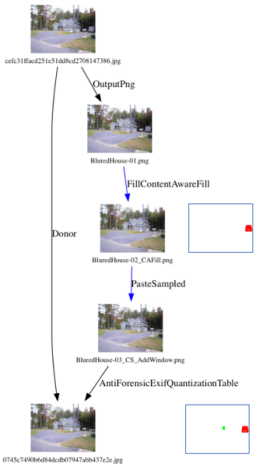
5.1. Image Manipulation Detection and Localization

The image Manipulation Detection and Localization (MDL) task is to detect if an image has been manipulated, and if so, to spatially localize the region determined to be manipulated. The reference ground-truth mask for localization is a JPEG 2000 image; each bit plane indicates the manipulated region(s) of a distinct manipulation operation step. Localizable manipulations (e.g., clone) have

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

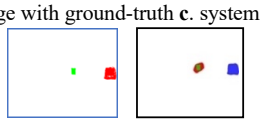
corresponding mask regions while global manipulations (e.g., blur) affecting the entire image do not. We also collect the following metadata: the camera information (camera ID, PRNU training data etc.); the image metadata (EXIF header information, captions, face(s) in image and their locations etc.); and the manipulation information (whether the image is manipulated, the major manipulation operations and their filters, parameters, and their orders, the corresponding masks for each operation step etc.). With rich metadata and manipulation information, we are able to provide a detailed evaluation report.

Figure 1 (a) shows a simple example of an image manipulation journal graph for localization evaluation. The top image in Figure 1 (a) is the nonmanipulated base image. There are two major manipulation steps: the manipulator removes the truck near the car using the content aware fill operation; and the manipulator clones a first-floor window.

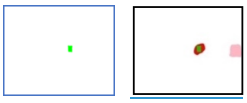


a. An example of image manipulation journal

b. test image with ground-truth **c**. system output results



d. full reference mask **e**. full eval. result: MCC = 0.196



f. selective scoring ref. **g**. selective eval.: MCC = 0.521

Figure 1: Image manipulation localization evaluation.

Two manipulated images could be used as testing images: the intermediate image after the truck is removed with only one major manipulation: content aware fill, and the final manipulated image with the truck removed (red region) and window cloned (green region). Given the final manipulated image as a test image, there are two types of evaluations. The first evaluation is on all manipulation operations regardless of manipulation type (Figure 1 (d) and (e)). The MCC of the system output mask shown in Figure 1 (c) is 0.196. The second evaluation is called selective localization (Figure 1 (f) and (g)); that is, we evaluate the system's performance on a single or a subset of manipulation operations only. If we evaluate the system performance only on the paste sample clone, then only the green region is used as the reference mask. Given the same system output as Figure 1 (c), the selective scoring MCC on paste sample operation is 0.521.

5.2. Video Detection and Temporal Localization

The video manipulation detection task is to detect if a video has been manipulated. Video temporal localization is to temporally localize the edit frame(s). The reference for temporal localization is the intervals where the frames were manipulated. Figure 2 shows the frames in a timeline, the blue part denotes the original frames, the green part denotes the manipulated frames. Given the ground-truth reference interval and the system detection result intervals, the evaluation scores can be reported given evaluation metrics.

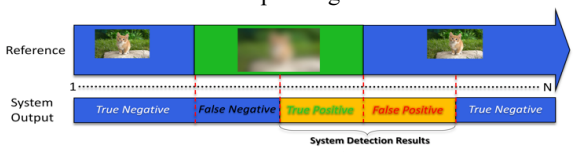


Figure 2: Video temporal localization.

5.3. Splice Manipulation Detection and Localization

Splice Detection (SD) and Localization (SDL) is to detect if a region of a given potential donor image has been spliced into a probe image and, if so, provide the mask regions for both images (Figure 3). Besides the data for the manipulated image described in MDL, the donor data such as the region in the donor image where the content was used for splice is also used for the evaluation.

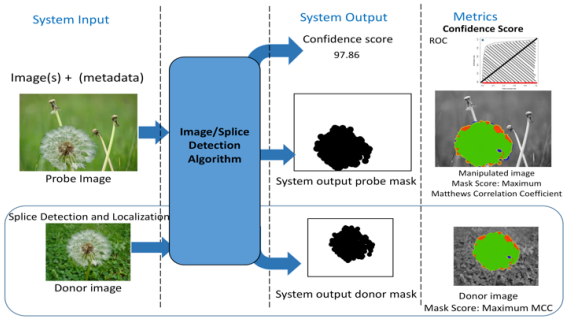


Figure 3: Splice detection and localization task.

5.4. Provenance Filtering and Graph Building

The Provenance Filtering (PF) task (Figure 4) is defined as follows: given a test image, find all images related to the test image from the given large-scale 'world' dataset. PF systems return up to n ancestors and descendants in the test image's phylogeny graph, including a base image and

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

donor images. The 'world' dataset contains a large portion of images downloaded from internet, a portion of nonmanipulated HP images, and all images from the manipulation journals. The test images are HP images and the manipulated images.

The Provenance Graph Building (PGB) task is to first retrieve related images with respect to the given query image from the world dataset, then reconstruct a provenance phylogeny graph; that is, the relationships among the associated images with manipulation sequences. There are two input dataset options for the graph building task: one is to use all data in the 'world' collection, the other is to use a small set of images, an 'oracle' set that contains all relevant images for the given probe with distractors. This oracle set allows the performer to do graph building without working on filtering first, testing PGB systems without being affected by the PF system's performance.

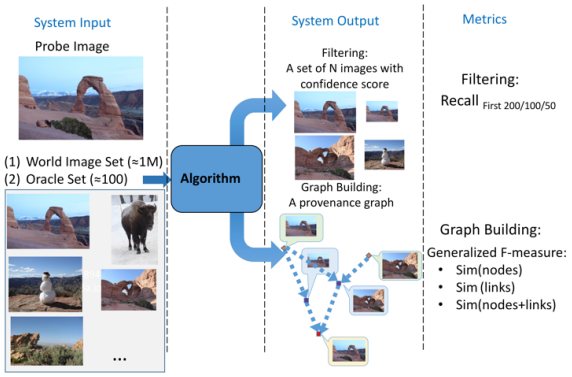


Figure 4: Provenance filtering and graph building task.

5.5. Event Verification

The MFC event verification task is defined as follows: given a collection of images (or videos) captured during an event (e.g. parade) and a probe image asserted to be captured during the event, verify if the probe image was taken during the event or if it was re-purposed. The data collection team collects images from several events such as air shows, hurricanes, marathons, blizzards etc. First, a small set of images is selected and released to the performer team as training images for each event, then another set of images is selected as testing images for each event. Each test image is then paired with each event name to generate the testing pairs serving as the performer's system input. The systems verify if the image belongs to the given event.

5.6. Camera Verification

The Camera Verification (CV) task is to verify if a media is captured by a claimed camera sensor. Distinct from the existing camera model detection task, which identifies the camera model given a media, CV focuses on sensor fingerprint verification. The traditional camera task trains and tests on the same modality (media type: image or video). We include cross-modality evaluations: e.g. the system trains on images, but tests on videos, which are captured from the same group of cameras. This gives the MFC18 CV task six data subsets, shown in Table 2. We also support the localization and selective scoring evaluation.

Table 2: MFC18 Camera Verification datasets.

Test	Train	Probe Pair	Target	Cam	Journal
Image (3761 img.)	Image	5275	2440	39	452
	Video	3383	1720	25	410
	Multimedia	3383	1720	25	410
Video (224 vid.)	Image	289	101	11	67
	Video	289	101	11	67
	Multimedia	289	101	11	67

5.7. GAN Challenge

GAN [1] are new technologies garnering a lot of recent attentions [46][47]. The GAN challenge is to evaluate if a system detects manipulated media produced by GAN-based manipulations. We created the MFC18 GAN full set (1340 images), GAN crop set (1000 images), and GAN video set (118 videos). The major image GAN operations include face swap, fill, erasure, camera model etc. The major video GAN operations include face swap, frame drop, erasure, and inpainting.

6. Evaluation Datasets

We have generated and released the following datasets: the Nimble Challenge 2016 (NC16) kickoff dataset, the NC17 development dataset, the NC17 evaluation dataset, the MFC18 development datasets 1 and 2, the MFC18 evaluation dataset, the MFC18 GAN image and video challenge datasets, the MFC18 camera ID verification dataset, and the MFC18 Event Verification development and evaluation datasets. All development datasets are publicly releasable. We partition the evaluation dataset into three subsets: Evaluation Part 1 (EP1) for public release and EP2 and EP3 for sequester evaluation.

Figure 5 shows the evolution of the evaluation datasets. The initial data collection, manipulation, and annotation started summer 2015. We designed, collected, manipulated, and released the NC16 kickoff dataset, which only contains single-step manipulation journals.

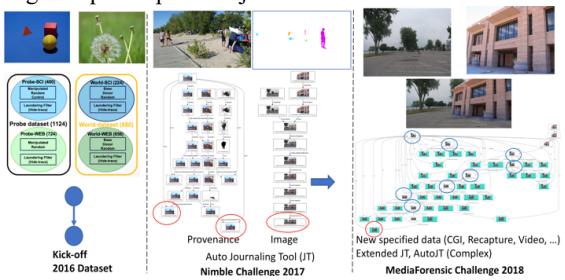


Figure 5: Media Forensic Challenge Dataset History.

In NC17, the data collection team joined the program; we designed and developed the JT and AutoJT, creating complex manipulations and collecting all data and metadata using manipulation journals. We built the NC17 development and evaluation datasets for the first year

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

MediFor evaluation. Later, the ExtendedJT was developed in 2018 to extend existing human journals with additional automatic manipulations to generate a large amount of manipulation testing images with little cost.

Table 3: MediFor Datasets summary.

Datasets	Image	Video	Size	Releasable
NC17 Dev	3.5 K	23	379	Y
NC17 EP1	4 K	45	3507	Y
NC17 Eval	6 K	1083	3600	N
MFC18 Dev 1	5.6 K	9	88.8	Y
MFC18 Dev 2	38 K	86	524	Y
MFC18 EP1	17 K	323	3200	Y
MFC18 Eval	80 K	2868	12300	N
MFC18 GAN	2.3 K	118	30	Y
MFC18 Cam	3.8 K	224	98	Y
MFC18 Event	2.4K	0	5.6	Y

Table 3 summarizes the number of images and videos, the total sizes, and the access permissions of our MFC development and evaluation datasets.

7. Generation of Evaluation Datasets

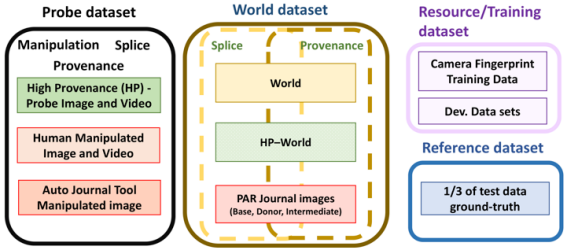


Figure 6: Media forensic evaluation dataset components.

To generate the testing data and its evaluation references, we developed an evaluation dataset generation infrastructure to build the dataset given raw data described in Section 4. Figure 6 shows the general components in the dataset for all tasks. Figure 7 shows the dataset production infrastructure.

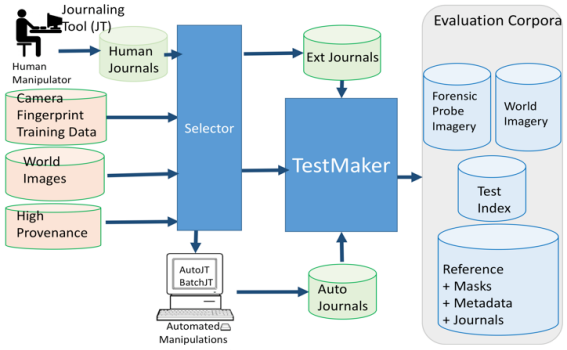


Figure 7: Evaluation datasets generation infrastructure.

We define the first camera set to control public releasable data. We release their PRNU images, development images, and the previous year's testing images to performers for PRNU training. We sequester another set of cameras and their images for future gradual release or sequester evaluation. To make sure the performer's system has enough data for training, we control the number of images released for each camera. Figure 8 shows how many images are released for each camera before the MFC18 evaluation.

Afterwards, we sample a set of manipulation journals for the dataset based on the manipulation operation distribution. Figure 9 shows a stacked histogram of the number of unique manipulation operations in different datasets. It shows that the NC17 datasets (yellow bar) have a limited number of manipulation operations: the distribution of the operations is uneven due to incorporating most of the journals available into the dataset. The MFC18 EP1 dataset's distribution covers most of the operations we wish to test. Note that some operation names were changed across datasets (e.g. "Intensity Normalization" in NC17 is covered by "Normalization" in MFC18). Finally, we select HP data and World data as nontarget for different tasks.

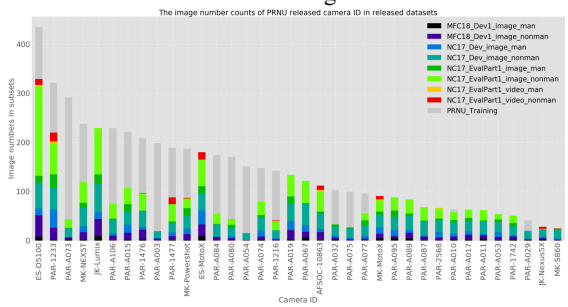


Figure 8: The stacked histogram of released image counts.

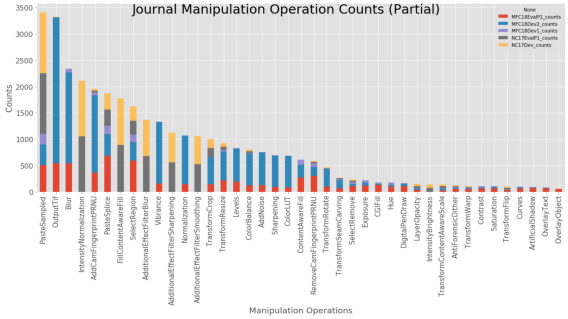


Figure 9: Stack histogram of manipulation operations.

Given the image/video manipulation journal, TestMaker extracts the data and metadata from the journal files, dynamically selects the testing images in the journals, and generates its JPEG2000 localization reference masks.

With the AutoJT and ExtendedJT, many testing images can be automatically generated with designated operations and parameters. Due to limited computational resources and evaluation time, we control the size of the test data by down sampling the manipulated images based on factor independence among data. Three approaches are used for data selection: (i) Random sampling, (ii) Specified sampling: e.g. traverse the longest journal path and select the middle and final node images. (iii) Sampling based on the distributions of the given factors.

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

8. MFC Evaluation Results

First, we demonstrate the selective scoring function to enable us to better understand the systems. Given the same set of systems, Figure 10 compares the evaluation results on all data with all manipulation types (left) with selective scoring results (right) on a particular manipulation (the target set is crop images extracted from the same test set). Two ROC curves (right) highlight the scoring system's ability to isolate the performance of components for particular manipulations. A separate document will discuss the MFC18 EP1 evaluation results in greater detail.

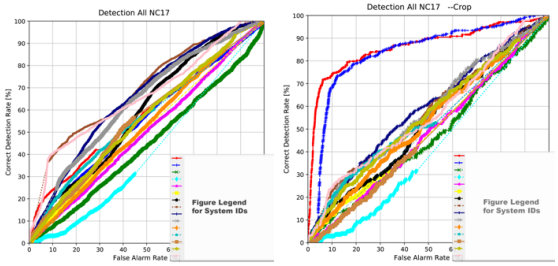


Figure 10: Full (left) vs. Selective Scoring on NC17 EP1.

Figure 11 shows an example graph building evaluation result. Green represents a correctly identified node/link, red represents an incorrect one, and gray represents a missing one. The graph node and link overlap similarity score (SimNLO, a generalized F-measurement) is 0.7.

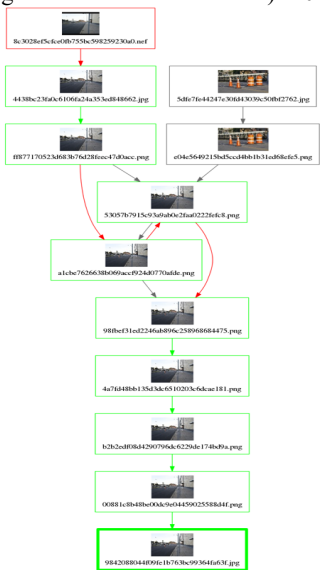


Figure 11: An example graph building evaluation.

Table 4 compares the performance of the best systems (the maximum score of all systems) of the NC17 and MFC18 evaluations. We compare the systems that evaluated all data of NC17 on NC17 EP1 with MFC18 on MFC18 EP1. The MFC18 EP1 and NC17 EP1 columns show: (i) the Max AUC for image detection systems improved from 0.69 to 0.84; (ii) the Max MCC for image localization systems improved from 0.08 to 0.26; (iii) the Max Recall for provenance filtering systems improved from 0.69 to 0.88; (iv) the Max SimNLO for provenance graph building systems on oracle systems on Full Graph metrics improved from 0.49 to 0.61. The system performance on NC17 Eval and NC17 EP1 is similar given the same conditions. After NC17 EP1 was released, the performances improved on both NC17 EP1 and Eval data.

Table 4: NC17 evaluation vs. MFC18 evaluation

Task and Metric Score	NC17 EP1	NC17 Eval	NC17 EP1 after NC17EP1 release	NC17 Eval after NC17EP1 release.	**MFC18** EP1
Image MD (AUC)	0.688	0.696	0.822	0.858	0.837
Image MDL (MCC)	0.082	0.009	0.290	-	0.258
PF (Recall)	0.689	0.649	0.782	0.688	0.882
PGB (SimNLO)	0.489	-	-	-	0.612
SD (AUC)	0.769	0.795	0.832	0.863	0.767
SDL (MCC)	-	-	0.295	-	0.361
Video MD (AUC)	-	0.580	-	0.580	0.594

9. Conclusion

We proposed an approach on designing and building an evaluation benchmark on a new research domain: media forensics. We present a series of datasets for different types of media forensic system evaluations and compare the evaluation results from the last two years' evaluations. We are continuing to collect and generate more data for MFC19 and MFC20 evaluations. The proposed methodology could be further generalized to other research domains.

The released datasets for NC17[2] and MFC18[3] are available upon request via email: mfc_poc@nist.gov. The evaluation scoring software package, MediScore, can be downloaded from https://github.com/usnistgov/MediScore.

10. Acknowledgement

The authors gratefully acknowledge the members of the DARPA Media Forensics (MediFor) Program and members of the Air Force Research Lab for managing the program and weekly data team meetings; special thanks goes to David Doermann and Rajiv Jain for their direct instructions and support. The authors would like to thank other PAR Government team members for media collection, manipulation, JT development; Rochester Institute of Technology, Drexel University, and University of Michigan members for their technical supports to PAR government; University of Colorado Denver for their professional media manipulations, antiforensic tools, social media laundering tools; the RankOne MediFor team for world data collection and the Medifor Browser; and Wendy Dinova-Wimmer for the NC16 Kickoff professional data manipulations and initial manipulation data collection design. PAR Government conducted this work under DARPA sponsorship via Air Force Research Laboratory (AFRL) contract FA8750-16-C-0168. NIST conducted this work under NIST Interagency Agreement Number 1505-774-08-000.

[2] https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation

[3] https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2018

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

References

[1] T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, International Conference on Learning Representations, 2018.

[2] T. Wang, M. Liu, J. Zhu, G. Liu, A. Tao, J. Kautz, B. Catanzaro, Video-to-Video Synthesis, arXiv:1808.06601, 2018

[3] A. Piva, An Overview on Image Forensics, ISRN Signal Processing, 2013:1–22, 2013.

[4] M. C. Stamm, M. Wu, and K. J. R. Liu, Information Forensics: An Overview of the First Decade, IEEE Access, 1:167–200, 2013.

[5] H. Farid, Image forgery detection, IEEE Signal Process. Mag., 26(2):16–25, 2009.

[6] J. Fridrich, Digital image forensics using sensor noise, Signal Proc. Mag. IEEE, 26(2): 26–37, 2009.

[7] H. T. Sencar and N. Memon, "Overview of state-of-the-art in digital image forensics," Algorithms Archit. Inf. Syst. Secur., 3:325–348, 2008.

[8] R. Boehme, M. Kirchner, Counter-Forensics: Attacking Image Forensics, Digital Image Forensics, H. T. Sencar and N. D. Memon, eds., pp. 327-366, Springer, 2012.

[9] Siwei Lyu, Natural Image Statistics in Digital Image Forensics, Springer, Digital Image Forensics, pp 239-256, 2012.

[10] S. Tariq, S. Lee, H. Kim, Y. Shin, S. S. Woo, Detecting Both Machine and Human Created Fake Face Images In the Wild, the 2nd International Workshop on Multimedia Privacy and Security, pp. 81-87, 2018.

[11] DARPA MediFor, https://www.darpa.mil/program/media-forensics,https://www.fbo.gov/index?s=opportunity&mode=form&id=bfa29e5f04566fbb961cd773a8a8649f&tab=core&_cview=1.

[12] A. Yates; H. Guan; Y. Lee, A. Delgado, D. Zhou, J. Fiscus, Nimble Challenge 2017 Evaluation Plan, NIST, 2017

[13] A. Yates; H. Guan; Y. Lee, A. Delgado, D. Zhou, J. Fiscus, Media Forensics Challenge 2018 Evaluation Plan, NIST, 2018

[14] G. Schaefer and M. Stich, UCID: an uncompressed color image database, Proceedings of the SPIE, 5307:472-480, 2003.

[15] J. Kalpathy-Cramer, A. Herrera, D. Demner-Fushman, S. Antani, S. Bedrick and H. Müller, Evaluating performance of biomedical image retrieval systems - an overview of the medical image retrieval task at ImageCLEF 2004-2013, Computerized Medical Imaging and Graphics, 39:55-61, 2015.

[16] G. Griffin, A. Holub, and P. Perona, Caltech-256 Object Category Dataset, California Institute of Technology. http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001, 2007.

[17] A. Torralba, R. Fergus, and W. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition, PAMI, 30(11):1958–1970, 2008.

[18] M. Fink and S. Ullman. From Aardvark to Zorro: A Benchmark for Mammal Image Classification. IJCV, 77(1-3):143–156, 2008.

[19] M. Everingham, S. M. Ali Eslami, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, The PASCAL Visual Object Classes Challenge: A Retrospective, International Journal of Computer Vision, 111(1): 98–136, 2015.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.

[21] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. SUN Database: Large-scale Scene Recognition from Abbey to Zoo, IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[22] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. International Journal of Robotics Research (IJRR), 2013.

[23] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, Microsoft COCO: Common Objects in Context, Computing Research Repository (CoRR), arXiv, abs/1405.0312, 2014.

[24] A. F. Smeaton, P. Over and W. Kraaij, Evaluation campaigns and TRECVid. MIR 2006 - 8th ACM SIGMM International Workshop on Multimedia Information Retrieval, 2006.

[25] S. Oh et al., A large-scale benchmark dataset for event recognition in surveillance video, CVPR, pp. 3153-3160, 2011.

[26] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, L. Li, YFCC100M: The New Data in Multimedia Research, Communications of the ACM, 59(2):64-73, 2016.

[27] M. Larson, M. Soleymani, G. Gravier, B. Ionescu and G. J. F. Jones, The Benchmarking Initiative for Multimedia Evaluation: MediaEval 2016, IEEE MultiMedia, 24(1):93-96, 2017.

[28] EU project REWIND (REVerse engineering of audio-VIsual coNtent Data), https://sites.google.com/site/rewindpolimi/home .

[29] A. Rocha, A. Piva, and J. Huang, The First IFS-TC Image Forensics Challenge, http://ifc.recod.ic.unicamp.br/.

[30] T.-T. Ng, S.-F. Chang, and Q. Sun, A data set of authentic and spliced image blocks, Columbia Univ. ADVENT Tech. Rep., pp. 203–2004, 2004.

[31] CalPhotos, A database of plants, animals, habitats and other natural history subjects, http://calphotos.berkeley.edu/, 2015.

[32] J. Dong, W. Wang, and T. Tan, CASIA image tampering detection evaluation database, in Signal and Information Processing (ChinaSIP), IEEE China Summit & International Conference on, pp. 422–426 2013.

[33] D. Tralic, I. Zupancic, S. Grgic and M. Grgic, CoMoFoD - New database for copy-move forgery detection, IEEE Proceedings ELMAR-2013, pp. 49-54, Zadar, 2013.

[34] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo, and G. Serra, A SIFT-based forensic method for copy-move attack detection and transformation recovery, IEEE Trans. Inf. Forensics Security, 6(3): 1099–1110, 2011.

[35] V. Christlein, C. Riess, J. Jordan, C. Riess, E. Angelopoulou, An Evaluation of Popular Copy-Move Forgery Detection Approaches, IEEE Transactions on Information Forensics and Security, 7(6):1841-1854, 2012.

[36] S. Agarwal, W. Fan, and H. Farid, A Diverse Large-Scale Dataset for Evaluating Rebroadcast Attacks, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1997-2001, 2018.

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

[37] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, R. Chellappa, UMDFaces: An Annotated Face Dataset for Training Deep Networks, arXiv:1611.01484v2,2016.
[38] D. Shullani, M. Fontani, M. Iuliani, O. Al Shaya, A. Piva, VISION: a video and image dataset for source identification, EURASIP Journal on Information Security, 2017:15, 2017.
[39] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces, arXiv:1803.09179, 2018.
[40] P. Bas, T. Filler, T. Pevny, "Break Our Steganographic System": The Ins and Outs of Organizing BOSS. INFORMATION HIDING, Czech Republic. 6958/2011:59-70, Lecture Notes in Computer Science, 2011.
[41] N. Khanna, A. K. Mikkilineni, G. T. Chiu, J. P. Allebach, E. J. Delp, Survey of Scanner and Printer Forensics at Purdue University. Computational Forensics. IWCF 2008. Lecture Notes in Computer Science, vol. 5158. Springer, 2008.
[42] M. Goljan, J. Fridrich, and T. Filler, Large scale test of sensor fingerprint camera identification, in Proc. SPIE Media Forensics and Security, Jan. 2009, vol. 7254.
[43] T. Gloe and R. Böhme, The Dresden Image Database for Benchmarking Digital Image Forensics, Proceedings of the 25th Symposium On Applied Computing (ACM SAC 2010), 2:1585–1591, 2010.
[44] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, G. Boato, RAISE – A Raw Images Dataset for Digital Image Forensics, ACM Multimedia Systems, 2015.
[45] E. Robertson, "Manual for MaskGen Journaling Tool", PAR Government Solution, 2018, online link: https://github.com/rwgdrummer/maskgen/blob/master/doc/MediForJournalingTool-public.pdf.
[46] Jennifer Finney Boylan, Will Deep-Fake Technology Destroy Democracy? The New York Times, Oct. 17, 2018. https://www.nytimes.com/2018/10/17/opinion/deep-fake-technology-democracy.html.
[47] Hilke Schellmann, Deepfake Videos Are Getting Real and That's a Problem, The Wall Street Journals, Oct. 15, 2018. https://www.wsj.com/articles/deepfake-videos-are-ruining-lives-is-democracy-next-1539595787.

Guan, Haiying; Kozak, Mark; Robertson, Eric; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation."
Paper presented at WACV 2019, Waikoloa, HI, United States. January 8, 2019 - January 10, 2019.

An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation

Justin A. Kauffman* and William L. George†
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Jonathan S. Pitt‡
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Fluid-structure interaction simulations where solid bodies undergo large deformations require special handling of the mesh motion for Arbitrarily Lagrangian-Eulerian (ALE) formulations. Such formulations are necessary when body-fitted meshes with certain characteristics, such as boundary layer resolution, are required to properly resolve the problem. We present an overset mesh method to accommodate such problems in which flexible bodies undergo large deformations, or where rigid translation modes of motion occur. To accommodate these motions of the bodies through the computational domain, an overset mesh enabled ALE formulation for fluid flow is discretized with the hybridizable discontinuous Galerkin (HDG) finite element method. The overset mesh framework applied to the HDG method enables the deforming and translating dynamic meshes to maintain quality without remeshing. Verification is performed to demonstrate that optimal order convergence $\mathcal{O}(k+1)$ is obtained for arbitrary overlap and approximation order k.

I. Introduction

Developing numerical methods that can efficiently and accurately model complex multiphysics problems, such as fluid-structure interactions (FSI) where the bodies experience large deformations is still an active area of research. In order to capture large changes in the computational domain, which include solid deformations and/or rigid translation modes of motion, it is necessary to first study an arbitrarily Lagrangian-Eulerian (ALE) formulation of the governing fluid equations. This particular building block is of critical importance because we will employ a monolithic solver for FSI simulations [1], which requires that the governing fluid equations and governing solid equations be cast in the same reference frame. Further, an ALE formulation automatically allows for dynamic bodies, and therefore dynamic meshes to be incorporated. This last point is the driving motivation for utilizing an overset mesh framework. We have previously developed an overset mesh framework with the mathematical discretization of the hybridizable discontinuous Galerkin (HDG) finite element method for static linear problems such as convection-diffusion and elastostatics [2]. Below we present motivation for the use of an overset mesh method and the HDG method.

As physical geometries increase in complexity and bodies are required to move relative to one another throughout the duration of a simulation, choices must be made on how to discretize the computational domain. One approach is to decompose the domain into smaller subdomains, which we can allow to overlap, that can focus on the complex features and can maintain the initial mesh quality even as bodies move relative to one another. This approach has additional complications such as: How to address cells from a background mesh that lie within a solid body? How do meshes communicate with each other, especially in a case when multiple meshes overlap at the same location? The primary focus of this work is to demonstrate our communication algorithm between overset meshes.

Before proceeding it is necessary to define some common nomenclature that we will use in this work. We refer to a *mesh* as a discretized representation of a physical domain into polygonal volumes (i.e., cells). An *overset mesh* is a collection of meshes that overlap to discretize the entire physical domain. An *overset mesh method* uses overlapping meshes to discretize the physical domain. The first overset mesh method was developed by researchers at the National Aeronatics and Space Administration in 1983 [3, 4] to perform simulations of bodies in relative motion. Overset mesh methods have been applied to a variety of applications [5–11] for fixed and dynamic mesh scenarios.

*Postdoctoral Associate, National Institute of Standards and Technology, 100 Bureau Drive, MS 8911, Gaithersburg, MD 20899
†Computational Scientist, National Institute of Standards and Technology, 100 Bureau Drive, MS 8911, Gaithersburg, MD 20899
‡Research Associate Professor, Hume Center for National Security and Technology, Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

Our c
a cell tha
shows tw
cell is the
are *donor*

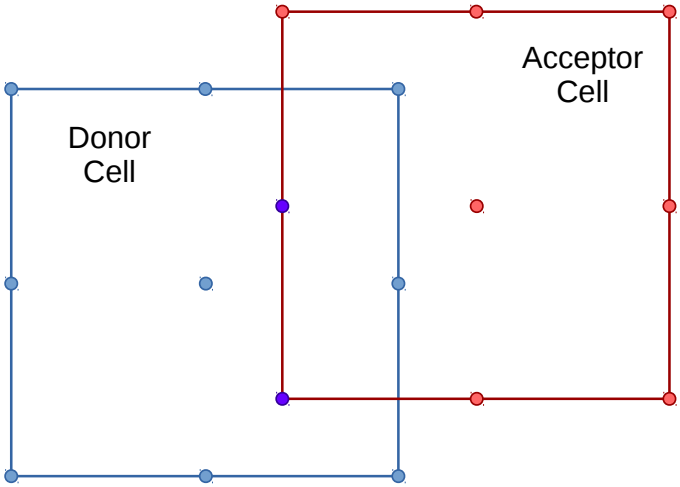


Fig. 1 The blue cell is a donor cell and the red cell is an acceptor cell. The purple points indicate where the basis functions of the blue cell will be evaluated to donate information to the red cell on that particular overset boundary. The points are distributed based on quadratic (Q2) discontinuous elements.

Higher-order finite difference and finite volume overset mesh methods require a larger overset region to maintain the higher-order approximations [12, 13]. This implies that more data must be stored, and also that the overset meshes must be manually constructed to ensure the correct amount of overlap during dynamic mesh motion applications is maintained. Nastase et al. [14] first remedied this problem by developing an overset discontinuous Galerkin (DG) approach for aerodynamic problems. The DG solver that replaced the finite difference solver is favorable because of the compact numerical stencil that results from this discretiztion. Similarly, Galbraith et al. [15] developed a DG Chimera scheme for solving the Euler and compressible Navier-Stokes equations. This scheme decomposes the computational domain into a set of structured overset meshes, which must at least abut to avoid gaps and voids in the domain. Communication via interpolation only occurs on the overset boundary faces, which is the same approach we are employing here and we first presented in [2]. The DG Chimera scheme also can handle mixed order meshes, so interpolation occurs between meshes of different approximation orders. Brazell et al. [16] also developed an overset DG approach within an overset mesh framework, `TIOGA` (Topology Independent Overset Grid Assembler). This approach performs interpolation over the entire overset boundary cell, instead of just the overset boundary face, which results in a different connectivity pattern compared to [15]. This DG-overset method investigates the Reynolds Averaged Naiver-Stokes equations with the Spalart-Allmaras turbulence model and utilizes unstructured mixed-element meshes. These DG-overset methods were the motivation to extend this concept to coupling overset meshes and the HDG finite element method.

The HDG method was first developed in [17] and provides an efficient way to reduce the size of the global system through the hybridization procedure. There have been previous HDG formulations for moving and deforming domains via an ALE perspective of the Navier-Stokes equations in Nguyen and Peraire [18] and Fidkowski [19]. The HDG ALE formulation has also been used to build an HDG FSI formulation in Sheldon et al. [1]. The HDG method has versatility and adaptability to enable straightforward implementation for a variety of different physics, which makes it ideal for complex multiphysics applications like FSI.

In the computational fluid dynamics (CFD) community DG methods are a viable alternative to finite volume methods, because this variant of the finite element method shares the conservation properties that are inherent to the finite volume method. Unfortunately, standard DG methods result in a linear system that becomes large due to duplication of degrees of freedom (DoFs) at nodes on element boundaries [18]. Through the hybridization process the global system is greatly reduced, when compared to standard DG methods, while maintaining the conservation properties that are necessary in

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

fluid dynamics. In Ahnert and Bärwolff [20] the authors compare an HDG method to a second-order finite volume method for incompressible flow; they conclude that the higher accuracy outweighs the slow matrix assembly for the HDG method compared to the finite volume method.

Hybridization involves introducing independent approximations on the border of elements (i.e., the trace of the elements) in addition to the interior of the elements for at least one field [21, Chapter 7]. Therefore, the new trace unknowns can be used to reduce the global system size and/or enrich the function space of the solutions. The HDG method decomposes the solution into two parts: the global solution trace, which exists on the cell boundaries, and the local solution, which exists internal to each cell. The local solution on each cell is only coupled with the global solution trace; therefore, the local solution on one cell is completely decoupled from the local solution of its neighboring cells. The only solution field that exists in both solution spaces is the hybrid unknown. The trace of the hybrid field approximates the global solution and is the only globally coupled field.

In this paper we are extending the overset mesh framework originally developed in [2] to an isentropic ALE Navier-Stokes HDG formulation. This formulation allows for bodies to move throughout the domain, relative to one another, and for the computational domain to deform. This paper serves as a proof-of-concept where we will verify our formulation using a four overset mesh configuration that demonstrates optimal order convergence is obtained for *arbitrary* overlap and *arbitrary* approximation order.

II. Governing Equations

In order to be incorporated in a monolithic solver the governing fluid equations are required to undergo a coordinate transformation from an Eulerian set of coordinates to an ALE set of coordinates. We describe this process below by first presenting the Eulerian form of the isentropic Navier-Stokes equations, where an equation of state for isentropic flow is

$$\frac{\partial \rho}{\partial t} = \beta \frac{\partial p}{\partial t}, \tag{1}$$

where $\beta = 1/c^2$ and c is the speed of sound through the fluid. Here we are neglecting the convective term from the material derivative and only relating the pure time derivatives, as in Bagwell [22]. Eq. (1) alters the continuity equation by incorporating time variance of pressure instead of density. The primary fields on the governing fluid system are velocity gradient $\mathbf{L}$, the velocity $\mathbf{v}$, and pressure p and the isentropic form of the governing Eulerian fluid equations over any physical domain Ω is

$$\mathbf{L} - \text{grad}\mathbf{v} = \mathbf{0}, \qquad \forall \boldsymbol{x} \in \Omega, \tag{2a}$$

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \text{div}\left(-\mu\mathbf{L} + p\mathbf{I}\right) + \rho\mathbf{L}\mathbf{v} + \rho\left(\text{div}\mathbf{v}\right)\mathbf{v} = \mathbf{f}, \qquad \forall \boldsymbol{x} \in \Omega, \tag{2b}$$

$$\varepsilon \frac{\partial p}{\partial t} + \text{div}\mathbf{v} = 0, \qquad \forall \boldsymbol{x} \in \Omega, \tag{2c}$$

$$\left(-\mu\mathbf{L} + p\mathbf{I}\right)\mathbf{n} = \mathbf{g}_N, \quad \forall \boldsymbol{x} \in \partial\Omega_N, \tag{2d}$$

$$\mathbf{v} = \mathbf{g}_D, \quad \forall \boldsymbol{x} \in \partial\Omega_D, \tag{2e}$$

where $\varepsilon = \beta/\rho$, $\mathbf{g}_N$ is a prescribed traction on $\partial\Omega_N$, and $\mathbf{g}_D$ is a prescribed velocity on $\partial\Omega_D$. In order to reformulate Eq. (2) to an ALE perspective we require the use of the mesh displacment $\mathbf{u}_\text{m}$ and the mesh deformation gradient $\mathbf{F}_\text{m}$ through the following relations

$$\mathbf{F}_\text{m} = \mathbf{I} + \text{Grad}\mathbf{u}_\text{m}, \tag{3a}$$

$$\text{grad}p = \mathbf{F}_\text{m}^{-\mathsf{T}}\text{Grad}p, \tag{3b}$$

$$\text{grad}\mathbf{v} = \left(\text{Grad}\mathbf{v}\right)\mathbf{F}_\text{m}^{-1}, \tag{3c}$$

$$\text{div}\mathbf{L} = \text{Grad}\mathbf{L} : \mathbf{F}_\text{m}^{-\mathsf{T}}, \tag{3d}$$

$$\text{div}\mathbf{v} = \text{Grad}\mathbf{v} : \mathbf{F}_\text{m}^{-\mathsf{T}}, \tag{3e}$$

where the subscript m indicates that the field corresponds to the governing equations of the mesh. Also, the : operator in Eqs. (3d) and (3e) is a double contraction [23] meaning that two indices are summed over and results in a scalar field if : acts between two second order tensors (as in Eq. (3e)), or in a vector if : acts between a third order tensor and a second order tensor (as in Eq. (3d)). We should note that grad and Grad are distinguished based on the coordinates,

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

and therefore the reference configuration, that these derivatives are being performed on, namely grad $:= \partial/\partial \boldsymbol{x}$ and Grad $:= \partial/\partial \mathbf{X}$, which is consistent with [24]. Additionally, we need the so-called 'fundamental ALE equation' [25] in both scalar and vector form:

$$\phi' = \dot{\phi} - \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathrm{Grad}\phi \cdot \frac{\partial \mathbf{X}}{\partial t}, \tag{4a}$$

$$\mathbf{a}' = \dot{\mathbf{a}} - \left[(\mathrm{Grad}\mathbf{a})\, \mathbf{F}_{\mathrm{m}}^{-1} \right] \frac{\partial \mathbf{X}}{\partial t}, \tag{4b}$$

where $\partial \mathbf{X}/\partial t = \mathbf{v}_{\mathrm{m}}$, the mesh velocity, the prime derivative is a material time derivative, and the dot derivative is an ordinary time derivative ($\partial/\partial t$). Lastly, normal vectors viewed from the ALE perspective instead of the Eulerian perspective undergo the following transformation

$$\mathbf{n} = \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}}, \tag{5}$$

where $\mathbf{n}$ is the original fluid normal vector in the Eulerian reference frame and $\mathbf{n}_{\mathrm{f}}$ is the fluid normal vector in the ALE reference frame. Using Eqs. (2)–(5) we can obtain the isentropic form of the governing ALE equations. This governing system contains two parts, the governing fluid equations and the governing mesh equations. The partial differential equations (PDE) system governing the fluid mechanics, which consists of primary fields velocity gradient $\mathbf{L}_{\mathrm{f}}$, velocity $\mathbf{v}_{\mathrm{f}}$, and pressure p_{f} is

$$\mathbf{L}_{\mathrm{f}} - \mathrm{Grad}\mathbf{v}_{\mathrm{f}} \mathbf{F}_{\mathrm{m}}^{-1} = \mathbf{0}, \qquad \forall \mathbf{X} \in \Omega_{\mathrm{F}}(0), \tag{6a}$$

$$\rho_{\mathrm{f}} \frac{\partial \mathbf{v}_{\mathrm{f}}}{\partial t} + \rho_{\mathrm{f}} \mathbf{L}_{\mathrm{f}} (\mathbf{v}_{\mathrm{f}} - \mathbf{v}_{\mathrm{m}}) - \mu_{\mathrm{f}} \mathrm{Grad}\mathbf{L}_{\mathrm{f}} : \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} + \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathrm{Grad} p_{\mathrm{f}} + \rho_{\mathrm{f}} \left(\mathrm{Grad}\mathbf{v}_{\mathrm{f}} : \mathbf{F}_{\mathrm{f}}^{-\mathsf{T}} \right) \mathbf{v}_{\mathrm{f}} = \mathbf{f}_{\mathrm{f}}, \qquad \forall \mathbf{X} \in \Omega_{\mathrm{F}}(0), \tag{6b}$$

$$\varepsilon \frac{\partial p_{\mathrm{f}}}{\partial t} - \varepsilon \left[\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathrm{Grad} p_{\mathrm{f}} \right] \cdot \mathbf{v}_{\mathrm{m}} + \mathrm{Grad}\mathbf{v}_{\mathrm{f}} : \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} = 0, \qquad \forall \mathbf{X} \in \Omega_{\mathrm{F}}(0), \tag{6c}$$

$$(-\mu_{\mathrm{f}} \mathbf{L}_{\mathrm{f}} + p_{\mathrm{f}} \mathbf{I}) \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}} = \mathbf{g}_{N_{\mathrm{f}}}, \quad \forall \mathbf{X} \in \partial\Omega_{N_{\mathrm{F}}}(0), \tag{6d}$$

$$\mathbf{v}_{\mathrm{f}} = \mathbf{g}_{D_{\mathrm{f}}}, \quad \forall \mathbf{X} \in \partial\Omega_{D_{\mathrm{F}}}(0), \tag{6e}$$

where $\Omega_{\mathrm{F}}(0)$ is the initial physical fluid domain from the ALE perspective. The fields that are required in order to couple the fluid and mesh together are the deformation gradient $\mathbf{F}_{\mathrm{m}}$ and the mesh velocity $\mathbf{v}_{\mathrm{m}}$. The mesh velocity is not a primary field in the mesh governing system, but it is approximated via a backward difference of the mesh displacement $\mathbf{u}_{\mathrm{m}}$ in time [26]. The mesh motion is governed by an elastostatics formulation; therefore, we do not consider elastic wave propagation and the mesh PDE system is time-independent, but a mesh velocity can be approximated through the change in displacement. The PDE system governing the mesh motion, which consists of primary fields displacement $\mathbf{u}_{\mathrm{m}}$ and deformation gradient $\mathbf{F}_{\mathrm{m}}$ is

$$\mathbf{F}_{\mathrm{m}} - \mathbf{I} - \mathrm{Grad}\mathbf{u}_{\mathrm{m}} = \mathbf{0}, \qquad \forall \mathbf{X} \in \Omega_{\mathrm{F}}(0), \tag{7a}$$

$$-\mathrm{Div}\left[\mathbb{C} (\mathrm{Sym}\mathbf{F}_{\mathrm{m}} - \mathbf{I}) \right] = \mathbf{b}_{\mathrm{m}}, \qquad \forall \mathbf{X} \in \Omega_{\mathrm{F}}(0), \tag{7b}$$

$$\mathbb{C} (\mathrm{Sym}\mathbf{F}_{\mathrm{m}} - \mathbf{I}) \mathbf{n}_{\mathrm{m}} = \mathbf{t}_{N_{\mathrm{m}}}, \quad \forall \mathbf{X} \in \partial\Omega_{N_{\mathrm{F}}}(0), \tag{7c}$$

$$\mathbf{u}_{\mathrm{m}} = \bar{\mathbf{u}}_{D_{\mathrm{m}}}, \quad \forall \mathbf{X} \in \partial\Omega_{D_{\mathrm{F}}}(0), \tag{7d}$$

where $\mathbb{C}$ is the fourth order elasticity tensor, $\mathrm{Sym}\mathbf{F}_{\mathrm{m}} = 1/2 \left(\mathbf{F}_{\mathrm{m}} + \mathbf{F}_{\mathrm{m}}^{\mathsf{T}} \right)$, and $\mathbf{b}_{\mathrm{m}}$ is a source term in the balance of linear momentum. In order to present the variational (or weak) forms of Systems (6) and (7) we need to describe the relationship between the differential volume/area elements as we change our reference frame from Eulerian to ALE. This relation is

$$dv = J_{\mathrm{m}}^{\mathrm{h}} dv_{\mathrm{m}}, \quad da = J_{\mathrm{m}}^{\mathrm{h}} da_{\mathrm{m}}, \tag{8}$$

where $J_{\mathrm{m}}^{\mathrm{h}} = \det \mathbf{F}_{\mathrm{m}}^{\mathrm{h}}$. Since the weak form is a set of integral equations the differentials used in those integrals need to undergo the changes defined in Eq. (8).

III. Overset Hybridizable Discontinuous Galerkin Method

We have already written Systems (6) and (7) in first-order form since it is required for the HDG method. Before discretizing the aforementioned governing PDE systems we need to introduce some notation that will be used throughout

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

Table 1 Notation utilized in the HDG finite element method.

Notation	Description
d	The spatial degree of the problem $\{d = 2, 3\}$.
$\mathscr{T}_\mathrm{h}$	The entire computational domain.
N	Total number of meshes used to discretely describe $\mathscr{T}_\mathrm{h}$.
i	The mesh id number.
$\mathcal{T}_\mathrm{h}^i$	The discretized i^{th} computational domain (or triangulation).
$\partial\mathcal{T}_\mathrm{h}^i$	The union of all faces of domain $\mathcal{T}_\mathrm{h}^i$.
K	An individual cell in $\mathcal{T}_\mathrm{h}^i$.
∂K	The boundary of an individual cell.
F	The face of a cell.
$\mathcal{F}_\mathrm{h}^i$	Set of all faces, for all cells in $\mathcal{T}_\mathrm{h}^i$.
$P_k(K)$	Space of polynomials of degree $\leq k$ over K.
$P_k(F)$	Space of polynomials of degree $\leq k$ over F.

the rest of this paper. First, Table 1 describes common definitions that we will use. Note that $\mathscr{T}_\mathrm{h} = \bigcup_{i=1}^{N} \mathcal{T}_\mathrm{h}^i$. We also must define the local and global inner products. The local inner products are over cells

$$(a,b)_K \coloneqq \int_K ab, \qquad \langle a,b\rangle_{\partial K} \coloneqq \int_{\partial K} ab, \qquad \text{for scalars,} \tag{9a}$$

$$(\mathbf{a},\mathbf{b})_K \coloneqq \int_K \mathbf{a}\cdot\mathbf{b}, \qquad \langle \mathbf{a},\mathbf{b}\rangle_{\partial K} \coloneqq \int_{\partial K} \mathbf{a}\cdot\mathbf{b}, \qquad \text{for vectors,} \tag{9b}$$

$$(\mathbf{A},\mathbf{B})_K \coloneqq \int_K \mathbf{A}:\mathbf{B}, \qquad \langle \mathbf{A},\mathbf{B}\rangle_{\partial K} \coloneqq \int_{\partial K} \mathbf{A}:\mathbf{B}, \qquad \text{for tensors,} \tag{9c}$$

whereas the global inner products are over the entire discretized domain

$$(a,b)_{\mathcal{T}_\mathrm{h}^i} \coloneqq \sum_K (a,b)_K, \qquad \langle a,b\rangle_{\partial\mathcal{T}_\mathrm{h}^i} \coloneqq \sum_{\partial K} (a,b)_{\partial K}, \qquad \text{for scalars,} \tag{10a}$$

$$(\mathbf{a},\mathbf{b})_{\mathcal{T}_\mathrm{h}^i} \coloneqq \sum_K (\mathbf{a},\mathbf{b})_K, \qquad \langle \mathbf{a},\mathbf{b}\rangle_{\partial\mathcal{T}_\mathrm{h}^i} \coloneqq \sum_{\partial K} (\mathbf{a},\mathbf{b})_{\partial K}, \qquad \text{for vectors,} \tag{10b}$$

$$(\mathbf{A},\mathbf{B})_{\mathcal{T}_\mathrm{h}^i} \coloneqq \sum_K (\mathbf{A},\mathbf{B})_K, \qquad \langle \mathbf{A},\mathbf{B}\rangle_{\partial\mathcal{T}_\mathrm{h}^i} \coloneqq \sum_{\partial K} (\mathbf{A},\mathbf{B})_{\partial K}, \qquad \text{for tensors.} \tag{10c}$$

Finally, we introduce the local discontinuous approximation spaces as

$$\mathscr{G}_\mathrm{h} \coloneqq \left\{ \mathbf{G}^\mathrm{h} \in \left[L^2\left(\mathcal{T}_\mathrm{h}^i\right)\right]^{d\times d} : \mathbf{G}^\mathrm{h}|_K \in [P_k(K)]^{d\times d},\ \forall K \in \mathcal{T}_\mathrm{h}^i,\ \forall i \right\}, \tag{11a}$$

$$\mathscr{Y}_\mathrm{h} \coloneqq \left\{ \mathbf{y}^\mathrm{h} \in \left[H^1\left(\mathcal{T}_\mathrm{h}^i\right)\right]^{d} : \mathbf{y}^\mathrm{h}|_K \in [P_k(K)]^{d},\ \forall K \in \mathcal{T}_\mathrm{h}^i,\ \forall i \right\}, \tag{11b}$$

$$\mathscr{P}_\mathrm{h} \coloneqq \left\{ q^\mathrm{h} \in L^2\left(\mathcal{T}_\mathrm{h}^i\right) : q^\mathrm{h}|_K \in P_k(K),\ \forall K \in \mathcal{T}_\mathrm{h}^i,\ \forall i \right\}, \tag{11c}$$

$$\mathscr{C}_\mathrm{h} \coloneqq \left\{ \mathbf{C}^\mathrm{h} \in \left[L^2\left(\mathcal{T}_\mathrm{h}^i\right)\right]^{d\times d} : \mathbf{C}^\mathrm{h}|_K \in [P_k(K)]^{d\times d},\ \forall K \in \mathcal{T}_\mathrm{h}^i,\ \forall i \right\}, \tag{11d}$$

$$\mathscr{U}_\mathrm{h} \coloneqq \left\{ \mathbf{w}^\mathrm{h} \in \left[H^1\left(\mathcal{T}_\mathrm{h}^i\right)\right]^{d} : \mathbf{w}^\mathrm{h}|_K \in [P_k(K)]^{d},\ \forall K \in \mathcal{T}_\mathrm{h}^i,\ \forall i \right\}, \tag{11e}$$

where the L^2 is a Hilbert space where functions are square Lebesgue integrable and H^1 is a Sobolev space where the first order weak derivatives of a function exists. Additionally, we also have to introduce two approximation spaces for

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

the global fields, i.e., the trace of the solutions. Each system requires its own solution trace (velocity for the fluid system and displacement for the mesh system). The discontinuous approximation spaces for the global fields are

$$\mathscr{N}_{\mathrm{h}} \coloneqq \left\{ \eta^{\mathrm{h}} \in \left[L^2\left(\mathcal{F}_{\mathrm{h}}^i\right)\right]^d : \eta^{\mathrm{h}}|_F \in [P_k(F)]^d,\ \forall F \in \mathcal{F}_{\mathrm{h}}^i,\ \forall K \in \mathcal{T}_{\mathrm{h}}^i,\ \forall i \right\}, \tag{12a}$$

$$\mathscr{M}_{\mathrm{h}} \coloneqq \left\{ \omega^{\mathrm{h}} \in \left[L^2\left(\mathcal{F}_{\mathrm{h}}^i\right)\right]^d : \omega^{\mathrm{h}}|_F \in [P_k(F)]^d,\ \forall F \in \mathcal{F}_{\mathrm{h}}^i,\ \forall K \in \mathcal{T}_{\mathrm{h}}^i,\ \forall i \right\}. \tag{12b}$$

A. Overset Framework

For an overset mesh framework the physical domain is decomposed into N subdomains that must at least abut, but will overlap in general. Mathematically, this decomposition is

$$\Omega = \Omega^1 \cup \Omega^2 \cup \cdots \cup \Omega^N \tag{13}$$

Domains are considered to overlap if

$$\Omega^i \cap \Omega^j \neq \emptyset, \quad \text{for any } i, j \in \{1, \ldots, N\}, \quad i \neq j, \tag{14}$$

and considered to abut if they share a common boundary, but do not otherwise overlap, namely,

$$\overline{\Omega}^i \cap \overline{\Omega}^j \neq \emptyset, \quad \Omega^i \cap \Omega^j = \emptyset, \quad \text{for any } i, j \in \{1, \ldots, N\}, \quad i \neq j. \tag{15}$$

Communication between the overset meshes is achieved via local (interior/volume) field information being donated from one domain to the face/edge of another domain on integrals over the overset boundaries (the purple points in Figure 1 where the interior information is being donated from the blue cell). The coupling between domains is a face-volume, or global-local, coupling which occurs naturally in the HDG method. Also, as we will show in Section III.B the local problem, which is solved on each cell K in each $\mathcal{T}_{\mathrm{h}}^i$, does not change. More details about the linear algebra are provided in Section III.C. Donated fields will be represented with a breve accent, $\breve{p}$ for example. This implies that the local field contributions to the integrals on the overset boundaries have information from the interior of a different mesh being included (i.e., donated).

B. Overset Formulation

The weak form of the isentropic ALE Navier-Stokes problem in an overset mesh framework is obtained by multiplying each equation in Systems (6) and (7) by a corresponding test function and then integrating over each cell in each domain. The test functions are $\{\mathbf{G}, \mathbf{y}, q\}$ for the fluid system and $\{\mathbf{C}, \mathbf{w}\}$ for the mesh system. We used integration by parts to obtain boundary terms on each element in order to define the numerical traces between cells. The local weak form is obtained by observing one cell $K \in \mathcal{T}_{\mathrm{h}}^i$ such that the following is satisfied:

Local Problem 1 (Local Isentropic ALE Navier-Stokes) *Find* $\left(\mathbf{L}_{\mathrm{f}}^{\mathrm{h}}, \mathbf{v}_{\mathrm{f}}^{\mathrm{h}}, p_{\mathrm{f}}^{\mathrm{h}}, \mathbf{F}_{\mathrm{m}}^{\mathrm{h}}, \mathbf{u}_{\mathrm{m}}^{\mathrm{h}}\right) \in \mathscr{G}_{\mathrm{h}} \times \mathscr{V}_{\mathrm{h}} \times \mathscr{P}_{\mathrm{h}} \times \mathscr{C}_{\mathrm{h}} \times \mathscr{U}_{\mathrm{h}}$ *such that*

Local Subproblem 1.1 (Fluid)

$$\left(\mathbf{G}, J_{\mathrm{m}}^{\mathrm{h}} \mathbf{L}_{\mathrm{f}}^{\mathrm{h}}\right)_K - \left(\mathbf{G}, J_{\mathrm{m}}^{\mathrm{h}} \mathrm{Grad}\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} \mathbf{F}_{\mathrm{m}}^{-1}\right)_K + \left\langle \mathbf{G}, J_{\mathrm{m}}^{\mathrm{h}} \left(\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} - \widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}\right) \otimes \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}} \right\rangle_{\partial K} = 0, \tag{16a}$$

$$\begin{aligned} \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}} \rho_{\mathrm{f}} \frac{\partial \mathbf{v}_{\mathrm{f}}^{\mathrm{h}}}{\partial t}\right)_K + \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}} \rho_{\mathrm{f}} \mathbf{L}_{\mathrm{f}}^{\mathrm{h}} \left[\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} - \mathbf{v}_{\mathrm{m}}^{\mathrm{h}}\right]\right)_K + \left(\mathrm{Grad}\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}} \left[\mu_{\mathrm{f}} \mathbf{L}_{\mathrm{f}}^{\mathrm{h}} - p_{\mathrm{f}}^{\mathrm{h}} \mathbf{I}\right] \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\right)_K \\ + \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}} \rho_{\mathrm{f}} \left[\mathrm{Grad}\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} : \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\right] \mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\right)_K + \left\langle \mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}} \widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}} \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}} \right\rangle_{\partial K} = \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}} \mathbf{f}_{\mathrm{f}}\right)_K, \end{aligned} \tag{16b}$$

$$\left(q, J_{\mathrm{m}}^{\mathrm{h}} \varepsilon \frac{\partial p_{\mathrm{f}}^{\mathrm{h}}}{\partial t}\right)_K + \left(q, J_{\mathrm{m}}^{\mathrm{h}} \varepsilon \left[\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathrm{Grad} p_{\mathrm{f}}^{\mathrm{h}}\right] \cdot \mathbf{v}_{\mathrm{m}}^{\mathrm{h}}\right)_K - \left(\mathrm{Grad}q, J_{\mathrm{m}}^{\mathrm{h}} \mathbf{F}_{\mathrm{m}}^{-1} \mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\right)_K + \left\langle q, J_{\mathrm{m}}^{\mathrm{h}} \widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}} \cdot \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}} \right\rangle_{\partial K} = 0, \tag{16c}$$

where

$$\widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}} \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}} \coloneqq \left[-\mu_{\mathrm{f}} \mathbf{L}_{\mathrm{f}}^{\mathrm{h}} + p_{\mathrm{f}}^{\mathrm{h}} \mathbf{I}\right] \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}} \mathbf{n}_{\mathrm{f}} + \mathbf{S}_{\mathrm{f}} \left(\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} - \widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}\right), \tag{17}$$

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

Local Subproblem 1.2 (Mesh Motion)

$$\left(\mathbf{C}, \mathbf{F}_{\mathrm{m}}^{\mathrm{h}} - \mathbf{I}\right)_K - \left(\mathbf{C}, \mathrm{Grad}\mathbf{u}_{\mathrm{m}}^{\mathrm{h}}\right)_K - \left\langle \mathbf{C}\mathbf{n}_{\mathrm{m}}, \left(\mathbf{u}_{\mathrm{m}}^{\mathrm{h}} - \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right)\right\rangle_{\partial K} = 0, \tag{18a}$$

$$\left(\mathrm{Grad}\mathbf{w}, \mathbb{C}\left[\mathrm{Sym}\mathbf{F}_{\mathrm{m}}^{\mathrm{h}} - \mathbf{I}\right]\right)_K - \left\langle \mathbf{w}, \widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}\right\rangle_{\partial K} = (\mathbf{w}, \mathbf{b}_{\mathrm{m}})_K, \tag{18b}$$

where

$$\widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}} \coloneqq \mathbb{C}\left(\mathrm{Sym}\mathbf{F}_{\mathrm{m}}^{\mathrm{h}} - \mathbf{I}\right)\mathbf{n}_{\mathrm{m}} - \mathbf{S}_{\mathrm{m}}\left(\mathbf{u}_{\mathrm{m}}^{\mathrm{h}} - \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right), \tag{19}$$

$\forall\, (\mathbf{G}, \mathbf{y}, q, \mathbf{C}, \mathbf{w}) \in \mathscr{G}_{\mathrm{h}} \times \mathscr{Y}_{\mathrm{h}} \times \mathscr{P}_{\mathrm{h}} \times \mathscr{C}_{\mathrm{h}} \times \mathscr{U}_{\mathrm{h}}$,

where $\mathbf{S}_{\mathrm{f}}$ and $\mathbf{S}_{\mathrm{m}}$ are second order stabilization tensors, which are defined as

$$\mathbf{S}_{\mathrm{f}} \coloneqq \left(2\mu_{\mathrm{f}} + \rho_{\mathrm{f}}\|\mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\|\right)\mathbf{I}, \tag{20a}$$

$$\mathbf{S}_{\mathrm{m}} \coloneqq 50\mu_{\mathrm{m}}\mathbf{I}, \tag{20b}$$

where $\|\cdot\|$ represents the standard L^2-norm, μ_{m} is the shear modulus of the linear elastic mesh motion, and that the factor of 50 in the definition of $\mathbf{S}_{\mathrm{m}}$ is for increased stability and convergence. The definition of $\mathbf{S}_{\mathrm{f}}$ is based off of the work by Sheldon et al. [1].

Local Problem 1 is then discretized in time using a third order backward difference formula (BDF3) [26]. This formula is

$$11y^n - 18y^{n-1} + 9y^{n-2} - 2y^{n-3} = 6\Delta t f\left(t^n\right). \tag{21}$$

The fully discrete local weak form is provided in the Appendix. After discretizing in time, it is possible to solve Local Problem 1, which is composed of Local Subproblem 1.1 and Local Subproblem 1.2, if the numerical traces $\widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}$ and $\widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}$ are known. To determine the solutions of the numerical traces we have to sum over all K, in each $\mathcal{T}_{\mathrm{h}}^i$, and enforce the boundary conditions and continuity of the numerical traces $\left\langle \mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial K}$ and $\left\langle \mathbf{w}, \widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}\right\rangle_{\partial K}$ where $\widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}$ and $\widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}$ are defined in Eq. (17) and (19). The resulting weak form is

Problem 1 (Global Isentropic ALE Navier-Stokes) *Find* $\left(\mathbf{L}_{\mathrm{f}}^{\mathrm{h}}, \mathbf{v}_{\mathrm{f}}^{\mathrm{h}}, p_{\mathrm{f}}^{\mathrm{h}}, \widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}, \mathbf{F}_{\mathrm{m}}^{\mathrm{h}}, \mathbf{u}_{\mathrm{m}}^{\mathrm{h}}, \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right) \in \mathscr{G}_{\mathrm{h}} \times \mathscr{Y}_{\mathrm{h}} \times \mathscr{P}_{\mathrm{h}} \times \mathscr{N}_{\mathrm{h}} \times \mathscr{C}_{\mathrm{h}} \times \mathscr{U}_{\mathrm{h}} \times \mathscr{M}_{\mathrm{h}}$ *such that*

Subproblem 1.1 (Fluid)

$$\left(\mathbf{G}, J_{\mathrm{m}}^{\mathrm{h}}\mathbf{L}_{\mathrm{f}}^{\mathrm{h}}\right)_{\mathscr{T}_{\mathrm{h}}} - \left(\mathbf{G}, J_{\mathrm{m}}^{\mathrm{h}}\mathrm{Grad}\mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-1}\right)_{\mathscr{T}_{\mathrm{h}}} + \left\langle \mathbf{G}, J_{\mathrm{m}}^{\mathrm{h}}\left(\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} - \widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}\right) \otimes \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}} = 0, \tag{22a}$$

$$\left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\rho_{\mathrm{f}}\frac{\partial \mathbf{v}_{\mathrm{f}}^{\mathrm{h}}}{\partial t}\right)_{\mathscr{T}_{\mathrm{h}}} + \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\rho_{\mathrm{f}}\mathbf{L}_{\mathrm{f}}^{\mathrm{h}}\left[\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} - \mathbf{v}_{\mathrm{m}}^{\mathrm{h}}\right]\right)_{\mathscr{T}_{\mathrm{h}}} + \left(\mathrm{Grad}\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\left[\mu_{\mathrm{f}}\mathbf{L}_{\mathrm{f}}^{\mathrm{h}} - p_{\mathrm{f}}^{\mathrm{h}}\mathbf{I}\right]\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\right)_{\mathscr{T}_{\mathrm{h}}}$$
$$+ \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\rho_{\mathrm{f}}\left[\mathrm{Grad}\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} : \mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\right]\mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\right)_{\mathscr{T}_{\mathrm{h}}} + \left\langle \mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}} - \left(\mathbf{y}, J_{\mathrm{m}}^{\mathrm{h}}\mathbf{f}_{\mathrm{f}}\right)_{\mathscr{T}_{\mathrm{h}}} = 0, \tag{22b}$$

$$\left(q, J_{\mathrm{m}}^{\mathrm{h}}\varepsilon\frac{\partial p_{\mathrm{f}}^{\mathrm{h}}}{\partial t}\right)_{\mathscr{T}_{\mathrm{h}}} + \left(q, J_{\mathrm{m}}^{\mathrm{h}}\varepsilon\left[\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathrm{Grad}p_{\mathrm{f}}^{\mathrm{h}}\right]\cdot\mathbf{v}_{\mathrm{m}}^{\mathrm{h}}\right)_{\mathscr{T}_{\mathrm{h}}} - \left(\mathrm{Grad}q, J_{\mathrm{m}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-1}\mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\right)_{\mathscr{T}_{\mathrm{h}}} + \left\langle q, J_{\mathrm{m}}^{\mathrm{h}}\widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}\cdot\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}} = 0, \tag{22c}$$

$$\left\langle \eta, J_{\mathrm{m}}^{\mathrm{h}}\widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}\setminus\partial\Omega_{D_{\mathrm{F}}}} \left\langle \eta, J_{\mathrm{m}}^{\mathrm{h}}\breve{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial\Omega_{O_{\mathrm{F}}}} - \left\langle \eta, J_{\mathrm{m}}^{\mathrm{h}}\mathbf{g}_N\right\rangle_{\partial\Omega_{N_{\mathrm{F}}}} = 0, \tag{22d}$$

$$\left\langle \eta, J_{\mathrm{m}}^{\mathrm{h}}\widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}\right\rangle_{\partial\Omega_{D_{\mathrm{F}}}} - \left\langle \eta, J_{\mathrm{m}}^{\mathrm{h}}\mathbf{g}_{D_{\mathrm{F}}}\right\rangle_{\partial\Omega_{D_{\mathrm{F}}}} = 0, \tag{22e}$$

where $\widehat{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}$ *is defined in Eq.* (17) *and* $\breve{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}$ *is defined as*

$$\breve{\mathbf{T}}_{\mathrm{f}}^{\mathrm{h}}\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}} \coloneqq \left[-\mu_{\mathrm{f}}\breve{\mathbf{L}}_{\mathrm{f}}^{\mathrm{h}} + \breve{p}_{\mathrm{f}}^{\mathrm{h}}\mathbf{I}\right]\mathbf{F}_{\mathrm{m}}^{-\mathsf{T}}\breve{\mathbf{n}}_{\mathrm{f}} + \mathbf{S}_{\mathrm{f}}\left(\breve{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}} - \widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}}\right). \tag{23}$$

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

Subproblem 1.2 (Mesh Motion)

$$\left(\mathbf{C}, \mathbf{F}_{\mathrm{m}}^{\mathrm{h}} - \mathbf{I}\right)_{\mathscr{T}_{\mathrm{h}}} - \left(\mathbf{C}, \mathrm{Grad}\mathbf{u}_{\mathrm{m}}^{\mathrm{h}}\right)_{\mathscr{T}_{\mathrm{h}}} - \left\langle \mathbf{C}\mathbf{n}_{\mathrm{m}}, \left(\mathbf{u}_{\mathrm{m}}^{\mathrm{h}} - \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right)\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}} = 0, \tag{24a}$$

$$\left(\mathrm{Grad}\mathbf{w}, \mathbb{C}\left[\mathrm{Sym}\mathbf{F}_{\mathrm{m}}^{\mathrm{h}} - \mathbf{I}\right]\right)_{\mathscr{T}_{\mathrm{h}}} - \left\langle \mathbf{w}, \widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}} = (\mathbf{w}, \mathbf{b}_{\mathrm{m}})_{\mathscr{T}_{\mathrm{h}}}, \tag{24b}$$

$$\left\langle \omega, \widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}\right\rangle_{\partial\mathscr{T}_{\mathrm{h}}\setminus\partial\Omega_{D_{\mathrm{F}}}} + \left\langle \omega, \breve{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}\right\rangle_{\partial\Omega_{O_{\mathrm{F}}}} = \left\langle \omega, \mathbf{t}_{N_{\mathrm{m}}}\right\rangle_{\partial\Omega_{N_{\mathrm{F}}}}, \tag{24c}$$

$$\left\langle \omega, \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right\rangle_{\partial\Omega_{D_{\mathrm{F}}}} = \left\langle \omega, \bar{\mathbf{u}}_{D_{\mathrm{m}}}\right\rangle_{\partial\Omega_{D_{\mathrm{F}}}}, \tag{24d}$$

where $\widehat{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}$ *is defined in Eq.* (19) *and* $\breve{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}}$ *is defined as*

$$\breve{\mathbf{P}}_{\mathrm{m}}^{\mathrm{h}}\mathbf{n}_{\mathrm{m}} := \mathbb{C}\left(\mathrm{Sym}\breve{\mathbf{F}}_{\mathrm{m}}^{\mathrm{h}} - \mathbf{I}\right)\breve{\mathbf{n}}_{\mathrm{m}} - \mathbf{S}_{\mathrm{m}}\left(\breve{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}} - \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right). \tag{25}$$

$\forall\, (\mathbf{G}, \mathbf{y}, q, \eta, \mathbf{C}, \mathbf{w}, \omega) \in \mathscr{G}_{\mathrm{h}} \times \mathscr{Y}_{\mathrm{h}} \times \mathscr{P}_{\mathrm{h}} \times \mathscr{N}_{\mathrm{h}} \times \mathscr{C}_{\mathrm{h}} \times \mathscr{U}_{\mathrm{h}} \times \mathscr{M}_{\mathrm{h}}$,

where $\partial\Omega_{D_{\mathrm{F}}} = \bigcup_{i=1}^{N} \partial\Omega_{D_{\mathrm{F}}}^{i}$, $\partial\Omega_{N_{\mathrm{F}}} = \bigcup_{i=1}^{N} \partial\Omega_{N_{\mathrm{F}}}^{i}$, and $\partial\Omega_{O_{\mathrm{F}}} = \bigcup_{i=1}^{N} \partial\Omega_{O_{\mathrm{F}}}^{i}$. Eqs. (23) and (25) show that all the local fields are donated from a mesh that is different from the one where the corresponding integral is being computed.

C. Implementation

The solution of Problem 1 is achieved through the Newton-Raphson method [27]. This method linearly approximates the residual of the nonlinear system allowing for the solution to be updated in an iterative fashion. The resulting linear system for a general overset mesh configuration is of the form

$$\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \begin{Bmatrix} \Upsilon \\ \Lambda \end{Bmatrix} = \begin{Bmatrix} \mathcal{H} \\ \mathcal{Q} \end{Bmatrix}, \tag{26}$$

where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ are block matrices and Υ, Λ, $\mathcal{H}$, $\mathcal{Q}$ are block vectors at each Newton iteration. For example,

$$\mathcal{A} := \begin{bmatrix} A^1 & 0 & \cdots & 0 \\ 0 & A^2 & \cdots & \vdots \\ 0 & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & A^N \end{bmatrix}. \tag{27}$$

The matrix $\mathcal{A}$ is a block diagonal matrix and each A^i is also a block diagonal matrix for the i^{th} mesh because the coupling between the entries in A^i only occurs within each $K \in \mathcal{T}_{\mathrm{h}}^{i}$, but not across neighboring elements because of the discontinuous basis functions used. Each block is representative of different degree of freedom coupling for the HDG formulation, namely,

$$\begin{bmatrix} \mathrm{VVC} & \mathrm{VFC} \\ \mathrm{FVC} & \mathrm{FFC} \end{bmatrix} \begin{Bmatrix} \text{local solution} \\ \text{global solution} \end{Bmatrix} = \begin{Bmatrix} \text{local source} \\ \text{global source} \end{Bmatrix}, \tag{28}$$

where V is volume, F is face, and C is coupling. In Section III.A we discussed that the coupling between meshes is a FVC, which corresponds to the $\mathcal{C}$ block matrix. In addition to the overset coupling in the $\mathcal{C}$ matrix we also have some extra FFC, the $\mathcal{D}$ matrix, which can be seen in the integrals associated with Eqs. (23) and (25). This extra FFC does not help tie the meshes together, but instead is a byproduct of the chosen numerical flux that enforces that the hybrid fields, fluid velocity and mesh displacement, are continuous across element boundaries. Overall, the structure of the linear system does not change for an overset configuration compared to a single mesh configuration. This implies that we can use the same linear algebra techniques that is common for the HDG method. Through static condensation [28], the local solution is condensed out of the linear system via the Schur complement [29] in order to obtain a system in terms of the global unknowns only,

$$\left(\mathcal{D} - \mathcal{C}\mathcal{A}^{-1}\mathcal{B}\right)\Lambda = \mathcal{Q} - \mathcal{C}\mathcal{A}^{-1}\mathcal{H}. \tag{29}$$

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

Since $\mathcal{A}$ is block diagonal it is easily inveritable. Solving Eq. (29) provides us with the global solution $\Lambda = \left\{\widehat{\mathbf{v}}_{\mathrm{f}}^{\mathrm{h}} \quad \widehat{\mathbf{u}}_{\mathrm{m}}^{\mathrm{h}}\right\}^{\intercal}$. The global solution is then used to compute the local solution via

$$\Upsilon^i = \left(\mathcal{A}^i\right)^{-1}\left(\mathcal{H}^i - \mathcal{B}^i\Lambda^i\right), \tag{30}$$

on a cell-by-cell basis for each $K \in \mathcal{T}_{\mathrm{h}}^i$ for each mesh $i = 1, \ldots, N$. We have implemented the overset coupling algorithm that is described in more detail in Kauffman et al. [2], and have utilized a sparse-direct linear solver `UMFPACK` [30] to obtain the global solution through the `deal.II` finite element library [31].

IV. Computational Studies

Problem 1 is the full isentropic ALE Navier-Stokes formulation. In terms of overset meshes this formulation allows individual meshes to deform and meshes to move relative to one another. This implies that meshes surrounding a body do not need to deform as bodies move, which maintains the initial mesh quality. Two-dimensional verification, through the method of manufactured solutions [32], of Problem 1 shows optimal order convergence for varying degrees and arbitrary amount of overlap in a four overset mesh system for varying approximation orders.

A. Two-dimensional verification

For the two-dimensional manufactured solution [32] of Problem 1 the velocity field is chosen as

$$\mathbf{v}^e\left(\mathbf{X}\right) = \begin{pmatrix}\cos\left(\pi\left[x+y\right]\right)\sin\left(\pi\left[x-y\right]\right)\cos\left(\frac{\pi}{4}t\right)\\ \cos\left(\pi\left[x+y\right]\right)\sin\left(\pi\left[x-y\right]\right)\cos\left(\frac{\pi}{4}t\right)\end{pmatrix}, \tag{31}$$

the pressure is chosen as

$$p^e\left(\mathbf{X}\right) = \frac{\cos\left(\pi\left[x+y\right]\right)\sin\left(\pi\left[x-y\right]\right)\cos\left(\frac{\pi}{4}t\right)}{\varepsilon}, \tag{32}$$

and the displacement is chosen as

$$\mathbf{u}^e\left(\mathbf{X}\right) = \begin{pmatrix}-x - \sin\left(\frac{\pi}{4}\left[x-y\right]\right)\cos\left(\frac{\pi}{4}t\right)\\ -y + \cos\left(\frac{\pi}{4}\left[x+y\right]\right)\cos\left(\frac{\pi}{4}t\right)\end{pmatrix}. \tag{33}$$

The velocity gradient is defined as $\mathbf{L}^e := \mathrm{Grad}\mathbf{v}^e\mathbf{F}^{-1}$ from Eq. (6a) and the deformation gradient is defined as $\mathbf{F}^e := \mathrm{Grad}\mathbf{u}^e + \mathbf{I}$ from Eq. (7a). In this formulation, because the pressure is chosen arbitrarily it is necessary to compute a right-hand-side term in Eq. (6c). The right-hand-side terms are not presented here but can be calculated using a symbolic algebra package. Fig. 2 shows the computational domain under investigation for the ALE verification problem. Note that because of the choice of the displacement field, care must be taken so that the determinant of the deformation gradient is not zero anywhere in the domain. The values of the parameters $\{\varepsilon, \rho_{\mathrm{f}}, \mu_{\mathrm{f}}, \lambda_{\mathrm{m}}, \mu_{\mathrm{m}}, \Delta t\}$ are chosen to be $\left\{1\ \left[\mathrm{m}\cdot\mathrm{s}^2/\mathrm{kg}\right],\ 1\ \left[\mathrm{kg}/\mathrm{m}^3\right],\ 1/2\ [\mathrm{Pa}\cdot\mathrm{s}],\ 1\ [\mathrm{Pa}],\ 1/2\ [\mathrm{Pa}],\ 10^{-6}\ [\mathrm{s}]\right\}$. Fig. 3 shows the four overset mesh system that is used in the verification of Problem 1. Table 2 describes the mesh domains, initial cell size and corresponding color in Fig. 3.

Table 2 Domain definitions and initial minimum cell sizes for all four meshes used for the verification of Problem 1. The set of points given for Ω^1 – Ω^4 correspond to lower left and upper right corners respectively.

Mesh	Fig. 3 Color	Domain	h_0
Ω^1	Orange	$[0.5, 0.5] \times [1.1, 1.1]$	0.2121
Ω^2	Green	$[0.8, 0.8] \times [1.5, 1.5]$	0.2475
Ω^3	Blue	$[0.5, 0.7] \times [1.2, 1.5]$	0.2658
Ω^4	Red	$[0.9, 0.5] \times [1.5, 1]$	0.1953

Convergence rates are demonstrated using the L^2 error norm, defined as

$$\|\mathbf{a}\|_{L^2} := \sqrt{\int_\Omega \mathbf{a}\cdot\mathbf{a}\, d\Omega}, \tag{34}$$

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

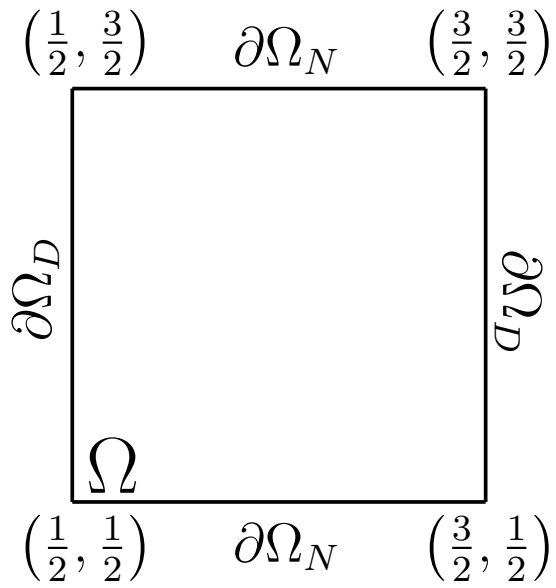


Fig. 2 Computational domain used for the two-dimensional verification of isentropic ALE Navier-Stokes with corresponding boundary conditions.

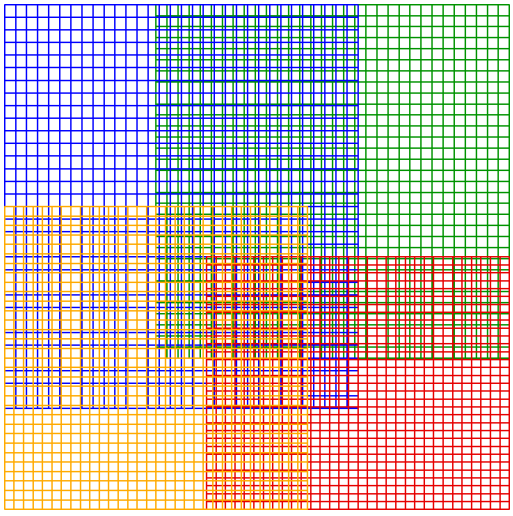


Fig. 3 The four overset mesh system used for the two-dimensional verification of isentropic ALE Navier-Stokes simulations.

for a generic vector field **a**. Similar definitions exist for scalar fields and tensor fields. Fig. 4 shows the convergence results for all of the linear elastic fields that govern the mesh motion: displacement and the deformation gradient. Both fields converge at the optimal rate for all approximation orders ($k = 1, \ldots, 4$) for each mesh. Fig. 5 shows the convergence results for all of the fluid fields: velocity, pressure, and velocity gradient. All fields converge at the optimal rate for all approximation orders ($k = 1, \ldots, 4$) for each mesh. The error in the velocity gradient for Ω^1 and Ω^4 starts to increase on the finest level mesh. This is attributed to the third order temporal scheme starting to dominate the error.

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

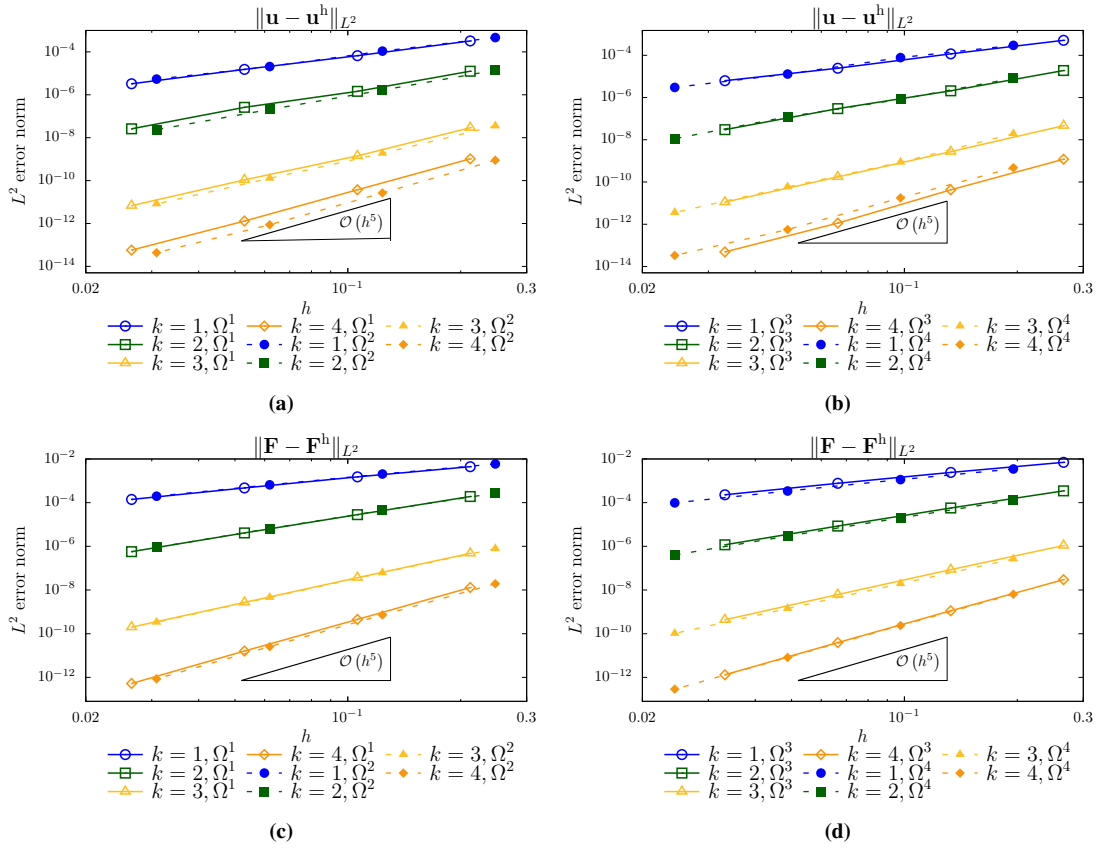


Fig. 4 Convergence results of the displacement u (a) and (b) and the deformation gradient F (c) and (d). All fields show optimal order convergence for all approximation orders. Note that this also verifies the overset system for the two-field linear elasticity problem, since no fluid properties affect this system.

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

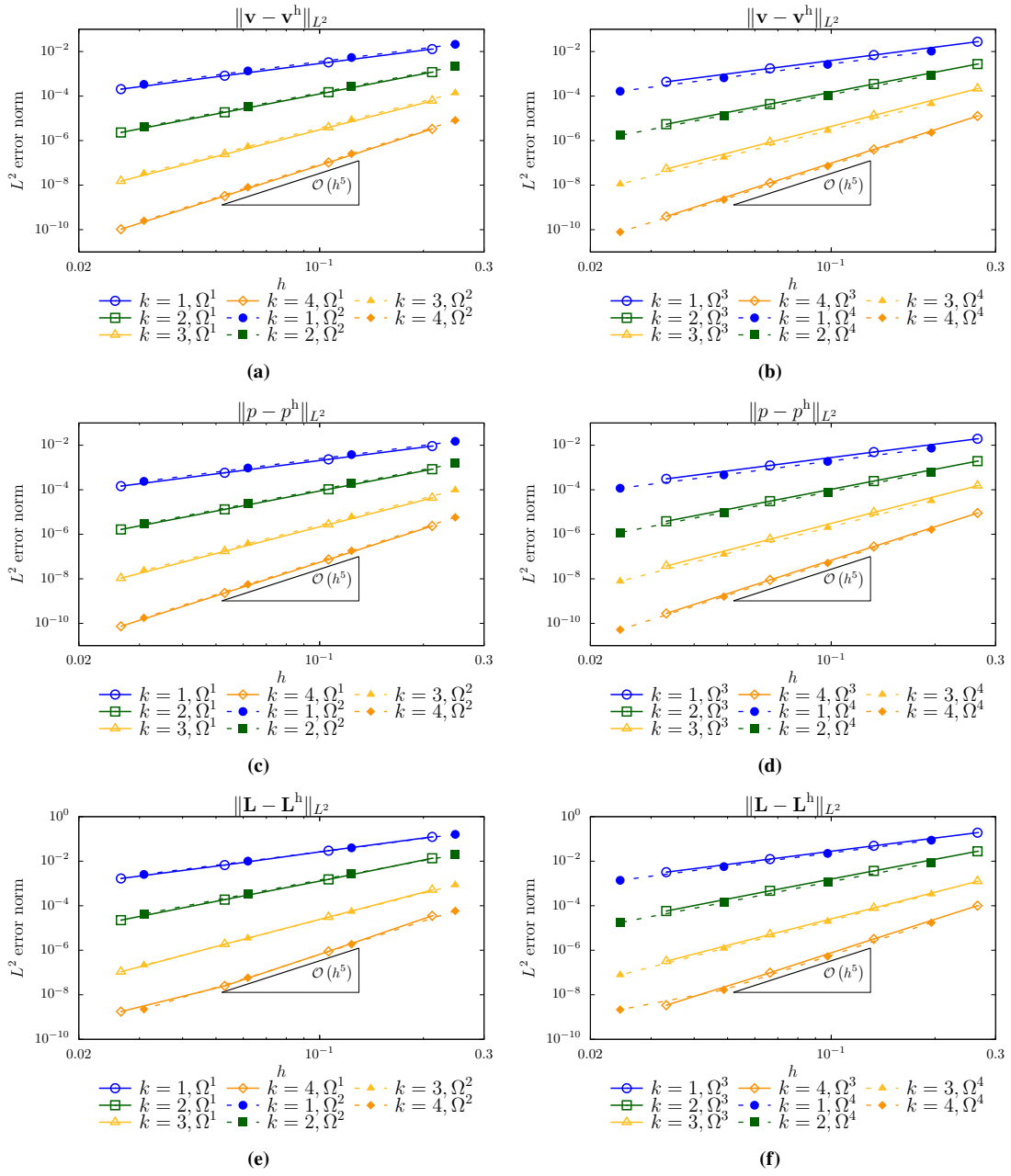


Fig. 5 Convergence results of the velocity v (a) and (b), the pressure p (c) and (d), and the velocity gradient L (e) and (f). All fields show optimal order convergence for all approximation orders. For $k = 4$ there is a slight deviation in the velocity gradient from the optimal slope. This is attributed to the temporal error dominating at the finest level mesh.

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

V. Summary

We have presented an overset-HDG method for an isentropic ALE Navier-Stokes formulation, which we have verified through the method of manufactured solutions in two-dimensions. Through verification, all fields, on each mesh, of the coupled system converged at optimal order $\mathcal{O}(k+1)$, where k is the approximation order. Through this verification procedure we have demonstrated that the amount of overlap between meshes is arbitrary and that convergence is optimal even for higher-order approximations.

VI. Future Work

The extensions of this work are to develop a parallel communication algorithm and implement a general block parallel solver. The parallelization of the communication algorithm will allow for more realistic simulations to be performed and direct comparisons between model experiments and other accepted overset methods. We expect that the general trend in the scaling between a single mesh configuration and a general N-mesh overset configuration will be similar once the communication algorithm is fully parallelized and a more robust linear solver is implemented. We are not expecting any overset configuration to run faster than a single mesh configuration. We will directly compare our fully parallel overset-HDG algorithm to existing overset codes already in use, and we expect that our algorithm will outperform the current overset standards.

Appendix

The fully discrete local weak formulation is presented below. This takes into account the third order backward difference formula (BDF3), defined in Eq. (21), used to discretize the Local Problem 1 in time. We also include the backward difference approximation we use to calculate the mesh velocity. For increased clarity the superscript h for all fields is removed and replaced with superscript n to indicate the current timestep. Note that $n-1$ refers to the previous timestep, and so on. So at timestep n the fully discrete weak form is

Local Problem 2 (Local Isentropic ALE Navier-Stokes) *Find* $\left(\mathbf{L}_{\mathrm{f}}^{n}, \mathbf{v}_{\mathrm{f}}^{n}, p_{\mathrm{f}}^{n}, \mathbf{F}_{\mathrm{m}}^{n}, \mathbf{u}_{\mathrm{m}}^{n}\right) \in \mathscr{G}_{\mathrm{h}} \times \mathscr{V}_{\mathrm{h}} \times \mathscr{P}_{\mathrm{h}} \times \mathscr{C}_{\mathrm{h}} \times \mathscr{U}_{\mathrm{h}}$ *such that*

Local Subproblem 2.1 (Fluid)

$$\left(\mathbf{G}, J_{\mathrm{m}}^{n}\mathbf{L}_{\mathrm{f}}^{n}\right)_{K} - \left(\mathbf{G}, J_{\mathrm{m}}^{n}\mathrm{Grad}\mathbf{v}_{\mathrm{f}}^{n}\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-1}\right)_{K} + \left\langle \mathbf{G}, J_{\mathrm{m}}^{n}\left(\mathbf{v}_{\mathrm{f}}^{n} - \widehat{\mathbf{v}}_{\mathrm{f}}^{n}\right) \otimes \left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial K} = 0, \tag{35a}$$

$$\begin{aligned}\left(\mathbf{y}, \frac{J_{\mathrm{m}}^{n}\rho_{\mathrm{f}}}{6\Delta t}\left[11\mathbf{v}_{\mathrm{f}}^{n} - 18\mathbf{v}_{\mathrm{f}}^{n-1} + 9\mathbf{v}_{\mathrm{f}}^{n-2} - 2\mathbf{v}_{\mathrm{f}}^{n-3}\right]\right)_{K} + \left(\mathbf{y}, J_{\mathrm{m}}^{n}\rho_{\mathrm{f}}\mathbf{L}_{\mathrm{f}}^{n}\left[\mathbf{v}_{\mathrm{f}}^{\mathrm{h}} - \left\{\frac{\mathbf{u}_{\mathrm{m}}^{n} - \mathbf{u}_{\mathrm{m}}^{n-1}}{\Delta t}\right\}\right]\right)_{K} \\ + \left(\mathrm{Grad}\mathbf{y}, J_{\mathrm{m}}^{n}\left[\mu_{\mathrm{f}}\mathbf{L}_{\mathrm{f}}^{n} - p_{\mathrm{f}}^{n}\mathbf{I}\right]\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\right)_{K} + \left(\mathbf{y}, J_{\mathrm{m}}^{n}\rho_{\mathrm{f}}\left[\mathrm{Grad}\mathbf{v}_{\mathrm{f}}^{n} : \left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\right]\mathbf{v}_{\mathrm{f}}^{n}\right)_{K} \\ + \left\langle \mathbf{y}, J_{\mathrm{m}}^{n}\widehat{\mathbf{T}}_{\mathrm{f}}^{n}\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial K} = \left(\mathbf{y}, J_{\mathrm{m}}^{n}\mathbf{f}_{\mathrm{f}}^{n}\right)_{K},\end{aligned} \tag{35b}$$

$$\begin{aligned}\left(q, \frac{J_{\mathrm{m}}^{n}\varepsilon}{6\Delta t}\left[11p_{\mathrm{f}}^{n} - 18p_{\mathrm{f}}^{n-1} + 9p_{\mathrm{f}}^{n-2} - 2p_{\mathrm{f}}^{n-3}\right]\right)_{K} + \left(q, J_{\mathrm{m}}^{n}\varepsilon\left\{\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\mathrm{Grad}p_{\mathrm{f}}^{n}\right\} \cdot \left\{\frac{\mathbf{u}_{\mathrm{m}}^{n} - \mathbf{u}_{\mathrm{m}}^{n-1}}{\Delta t}\right\}\right)_{K} \\ - \left(\mathrm{Grad}q, J_{\mathrm{m}}^{n}\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-1}\mathbf{v}_{\mathrm{f}}^{\mathrm{h}}\right)_{K} + \left\langle q, J_{\mathrm{m}}^{n}\widehat{\mathbf{v}}_{\mathrm{f}}^{n} \cdot \left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}}\right\rangle_{\partial K} = 0,\end{aligned} \tag{35c}$$

where

$$\widehat{\mathbf{T}}_{\mathrm{f}}^{n}\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}} := \left[-\mu_{\mathrm{f}}\mathbf{L}_{\mathrm{f}}^{n} + p_{\mathrm{f}}^{n}\mathbf{I}\right]\left[\mathbf{F}_{\mathrm{m}}^{n}\right]^{-\mathsf{T}}\mathbf{n}_{\mathrm{f}} + \mathbf{S}_{\mathrm{f}}^{n}\left(\mathbf{v}_{\mathrm{f}}^{n} - \widehat{\mathbf{v}}_{\mathrm{f}}^{n}\right), \tag{36}$$

Local Subproblem 2.2 (Mesh Motion)

$$\left(\mathbf{C}, \mathbf{F}_{\mathrm{m}}^{n} - \mathbf{I}\right)_{K} - \left(\mathbf{C}, \mathrm{Grad}\mathbf{u}_{\mathrm{m}}^{n}\right)_{K} - \left\langle \mathbf{C}\mathbf{n}_{\mathrm{m}}, \left(\mathbf{u}_{\mathrm{m}}^{n} - \widehat{\mathbf{u}}_{\mathrm{m}}^{n}\right)\right\rangle_{\partial K} = 0, \tag{37a}$$

$$\left(\mathrm{Grad}\mathbf{w}, \mathbb{C}\left[\mathrm{Sym}\mathbf{F}_{\mathrm{m}}^{n} - \mathbf{I}\right]\right)_{K} - \left\langle \mathbf{w}, \widehat{\mathbf{P}}_{\mathrm{m}}^{n}\mathbf{n}_{\mathrm{m}}\right\rangle_{\partial K} = \left(\mathbf{w}, \mathbf{b}_{\mathrm{m}}^{n}\right)_{K}, \tag{37b}$$

where

$$\widehat{\mathbf{P}}_{\mathrm{m}}^{n}\mathbf{n}_{\mathrm{m}} := \mathbb{C}\left(\mathrm{Sym}\mathbf{F}_{\mathrm{m}}^{n} - \mathbf{I}\right)\mathbf{n}_{\mathrm{m}} - \mathbf{S}_{\mathrm{m}}\left(\mathbf{u}_{\mathrm{m}}^{n} - \widehat{\mathbf{u}}_{\mathrm{m}}^{n}\right), \tag{38}$$

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

$\forall\, (\mathbf{G}, \mathbf{y}, q, \mathbf{C}, \mathbf{w}) \in \mathscr{G}_\mathrm{h} \times \mathscr{Y}_\mathrm{h} \times \mathscr{P}_\mathrm{h} \times \mathscr{C}_\mathrm{h} \times \mathscr{U}_\mathrm{h}$,

where $\mathbf{S}_\mathrm{f}^n$ is defined as

$$\mathbf{S}_\mathrm{f}^n := \left(2\mu_\mathrm{f} + \rho_\mathrm{f}\|\mathbf{v}_\mathrm{f}^{n-1}\|\right)\mathbf{I}. \tag{39}$$

Acknowledgments

J. Kauffman gratefully acknowledges financial support from the Applied Research Laboratory Walker Graduate Assistantship and the National Research Council postdoctoral fellowship.

References

[1] Sheldon, J. P., Miller, S. T., and Pitt, J. S., "A hybridizable discontinuous Galerkin method for modeling fluid–structure interaction," *Journal of Computational Physics*, Vol. 326, 2016, pp. 91–114.

[2] Kauffman, J. A., Sheldon, J. P., and Miller, S. T., "Overset meshing coupled with hybridizable discontinuous Galerkin finite elements," *International Journal for Numerical Methods in Engineering*, 2017. URL `http://dx.doi.org/10.1002/nme.5512`.

[3] Chan, W. M., "Overset grid technology development at NASA Ames Research Center," *Computer & Fluids*, Vol. 38, 2009, pp. 496–503.

[4] Sherer, S. E., Visbal, M. R., and Galbraith, M. C., "Automated Preprocessing Tools for Use with a High–Order Overset–Grid Algorithm," 44^{th} *AIAA Aerospace Sciences Meeting and Exhibit*, Vol. Overset and Adaptive Grids, Reno, Nevada, 2006, pp. 13–30.

[5] Bodony, D. J., Zagaris, G., Reichert, A., and Zhang, Q., "Provably stable overset grid methods for computational aeroacoustics," *Journal of Sound and Vibration*, Vol. 330, 2011, pp. 4161–4179.

[6] Noack, R. W., "SUGGAR: a General Capability for Moving Body Overset Grid Assembly," 17^{th} *AIAA Computational Fluid Dynamics Conference*, Vol. Overset Grids, Toronto, Ontario, Canada, 2005, pp. 5117–5138.

[7] Noack, R. W., Boger, D. A., Kunz, R. F., and Carrica, P. M., "Suggar++: An Improved General Overset Grid Assembly Capability," 19^{th} *AIAA Computational Fluid Dynamics*, Vol. Overset Grid Methods and Grid Refinement and Deformation Schemes, San Antonio, TX, 2009, pp. 3992–4040.

[8] Abras, J. N., Lynch, C. E., and Smith, M. J., "Computational Fluid Dynamics–Computational Structural Dynamics Rotor Coupling Using an Unstructured Reynolds–Averaged Navier–Stokes Methodology," *Journal of the American Helicopter Society*, Vol. 57, 2012, pp. 1–14.

[9] Foster, N. F., and Noack, R. W., "High–Order Overset Interpolation within an OVERFLOW Solution," 50^{th} *AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Expostion*, Vol. High-order Methods on Unstructured Grids III, Nashville, TN, 2012, pp. 728–736.

[10] Clark, C. G., Lyons, D. G., and Neu, W. L., "Comparison of Single and Overset Grid Techniques for CFD Simulations of a Surface Effect Ship," *ASME 2014* 33^{rd} *International Conference on Ocean, Offshore and Arctic Engineering*, Vol. CFD and VIV, San Francisco, CA, 2014, pp. 24207–24214.

[11] Östman, A., Pakozdi, C., Sileo, L., Stansberg, C.-T., and de Carvalho e Silva, D. F., "A Fully Nonliner RANS–VOF Numerical Wavetank Applied in the Analysis of Green Water on FPSO in Waves," $ASME33^{rd}$ *International Conference on Ocean, Offshore and Arctic Engineering*, Vol. Ocean Engineering, San Francisco, CA, 2014, pp. 23927–23935.

[12] Field, P. L., and Neu, W. L., "Comparison of RANS and Potential Flow Force Comutations for one DOF Heave and Pitch of the ONR Tumblehome Hullform," *ASME 2013* 32^{nd} *International Conference on Ocean, Offshore and Arctic Engineering*, Vol. Marine Hydrodynamics, Nantes, France, 2013, pp. 10562–10572.

[13] Sengupta, T. K., Suman, V., and Singh, N., "Solving Navier–Stokes equation for flow past cylinders using single–block structured and overset grids," *Journal of Computational Physics*, Vol. 229, 2010, pp. 178–199.

[14] Nastase, C., Mavriplis, D., and Sitaraman, J., "An overset unstructured mesh discontinuous Galerkin approach for aerodynamic problems," *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, 2011, p. 195.

[15] Galbraith, M. C., Benek, J. A., Orkwis, P. D., and Turner, M. G., "A Discontinuous Galerkin Chimera scheme," *Computers & Fluids*, Vol. 98, 2014, pp. 27–53.

[16] Brazell, M. J., Sitaraman, J., and Mavriplis, D. J., "An overset mesh approach for 3D mixed element high-order discretizations," *Journal of Computational Physics*, Vol. 322, 2016, pp. 33–51.

[17] Cockburn, B., Gopalakrishnan, J., and Lazarov, R., "Unified Hybridization of Discontinuous Galerkin, Mixed, and Continuous Galerkin Methods for Second Order Elliptic Problems," *SIAM Journal on Numerical Analysis*, Vol. 47, 2009, pp. 1319–1365.

[18] Nguyen, N., and Peraire, J., "Hybridizable discontinuous Galerkin method for partial differential equations in continuum mechanics," *Journal of Computational Physics*, Vol. 231, 2012, pp. 5955–5988.

[19] Fidkowski, K. J., "A hybridized discontinuous Galerkin method on mapped deforming domains," *Computers & Fluids*, Vol. 139, 2016, pp. 80–91.

[20] Ahnert, T., and Bärwolff, G., "Numerical comparison of hybridizable discontinuous Galerkin and finite volume methods for incompressible flow," *Internaltional Journal for Numerical Methods in Flu*, Vol. 76, 2014, pp. 267–281.

[21] Ciarlet, P. G., "The finite element method for elliptic problems," *Classics in applied mathematics*, Vol. 40, 2002, pp. 1–511.

[22] Bagwell, T. G., "CFD Simulation of Flow Tones from Grazing Flow Past a Deep Cavity," *ASME International Mechanical Engineering Congress and Exposition*, American Society of Mechanical Engineers, 2006, pp. 105–114.

[23] Holzapfel, G. A., *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*, John Wiley & Sons, LTD, New York, 2000.

[24] Gurtin, M. E., Fried, E., and Anand, L., *The Mechanics and Thermodynamics of Continua*, Cambridge University Press, New York, 2010.

[25] Donea, J., Huerta, A., Ponthot, J., and Rodriguez-Ferran, A., "Arbitrary Lagrangian Eulerian methods," *Fundamentals in Encyclopedia of Computational Mechanics*, Vol. 1, edited by E. Stein, R. de Borst, and T. J. Hughes, Wiley, New York, 2004, Chap. 14, pp. 413–437.

[26] Süli, E., and Mayers, D. F., *An introduction to Numerical Analysis*, Cambridge university press, 2003.

[27] Quarteroni, A., Sacco, R., and Saleri, F., *Numerical Mathematics*, Springer, 2000.

[28] Felippa, C. A., *Introduction to finite element methods*, Department of Aerospace Engineering Sciences, University of Colorado at Boulder, 2015. Accessed: 2016-01-15.

[29] Zhang, F., *The Schur complement and its applications*, Vol. 4, Springer Science & Business Media, 2005.

[30] Davis, T., "Algorithm 832: UMFPACK V4. 3—an unsymmetric-pattern multifrontal method," *ACM Transactions on Mathematical Software (TOMS)*, Vol. 30, No. 2, 2004, pp. 196–199.

[31] Alzetta, G., Arndt, D., Bangerth, W., Boddu, V., Brands, B., Davydov, D., Gassmoeller, R., Heister, T., Heltai, L., Kormann, K., Kronbichler, M., Maier, M., Pelteret, J., Turcksin, B., and Wells, D., "The `deal.II` Library, Version 9.0," *Journal of Numerical Mathematics*, 2018, accepted.

[32] Salari, K., and Knupp, P., "Code verification by the method of manufactured solutions," Tech. rep., Sandia National Labs., Albuquerque, NM (US); Sandia National Labs., Livermore, CA (US), 2000.

Kauffman, Justin; George, William; Pitt, Jonathan.
"An Overset Mesh Framework for an Isentropic ALE Navier-Stokes HDG Formulation."
Paper presented at AIAA Scitech 2019 Forum, San Diego, CA, United States. January 7, 2019 - January 11, 2019.

A Method-Level Test Generation Framework for Debugging Big Data Applications

Huadong Feng, Jaganmohan Chandrasekaran, Yu Lei
Department of Computer Science & Engineering
The University of Texas at Arlington
Arlington, USA
{huadong.feng, jaganmohan.chandrasekaran}@mavs.uta.edu,
ylei@cse.uta.edu

Raghu Kacker, D. Richard Kuhn
Information Technology Lab
National Institute of Standards and Technology
Gaithersburg, USA
{raghu.kacker, d.khun}@nist.gov

***Abstract*—When a failure occurs in a big data application, debugging with the original dataset can be difficult due to the large amount of data being processed. This paper introduces a framework for effectively generating method-level tests to facilitate debugging of big data applications. This is achieved by running a big data application with the original dataset and by recording the inputs to a small number of method executions, which we refer to as method-level tests, that preserve certain code coverage, e.g., edge coverage. The size of each method-level test is further reduced if needed, while maintaining code coverage. When debugging, a developer could inspect the execution of these method-level tests, instead of the entire program execution with the original dataset. We applied the framework to seven algorithms in the WEKA tool. The initial results show that in many cases a small number of method-level tests are sufficient to preserve code coverage. Furthermore, these tests could kill between 57.58% to 91.43% of the mutants generated using a mutation testing tool. This suggests that the framework could significantly reduce the efforts required for debugging big data applications.**

Keywords—Testing; Unit Testing; Big Data Application Testing; Test Generation; Test Reduction; Debugging; Mutation Testing;

I. INTRODUCTION

Big data applications are software programs that process large amounts of data. Debugging big data applications can be complicated and time-consuming. This is due to the fact that inspecting the execution of a big data application often involves long execution time, a large number of method executions, and/or a large number of objects. For example, a classification algorithm, called *DecisionTable*, in the WEKA tool [12] takes more than two hours to execute the *Heterogeneity Activity Recognition Dataset* (HAR) from the UC Irvine (UCI) Machine Learning Repository [13]. During the execution, one of the *DecisionTable's* methods, named *updateStatsForClassifier*, is executed more than half a billion times. (This method has 66 lines of code, not including comments and spaces.) If there exists a fault in this method, it can be very difficult to locate this fault due to the large number of times this method is executed.

Some approaches have been proposed to reduce the effort required for testing and debugging big data applications at the system level [1, 2, 3, 4, 5]. For example, data mining and machine learning methods are used to reduce the size of the original dataset or generate synthetic datasets [3, 4] for the testing purpose. The reduced dataset using such methods are executed at the system level, which can still be time-consuming. Furthermore, these methods are not designed to reproduce the original failure. Debugging approaches such as delta debugging [8] can identify the minimum failure-inducing input at the system level, which can reduce the size of the input while preserving the failure triggered by the original dataset. However, delta debugging can be very expensive for big data applications. This is because it requires the input data be recursively split into smaller chunks, each of which has to be executed at the system level. For big data applications, there can be a large number of chunks and system-level execution of each chunk can be time-consuming.

Our approach consists of two major steps. In the first step, we re-execute the failing system-level execution to record method-level tests for suspicious method(s). The main idea is to evaluate each method execution based on a chosen coverage criterion. In this paper, we used edge coverage, edge-pair coverage and edge-set coverage based on the Control Flow Graph (CFG) [11]. Note that other coverage criteria, e.g., prime-path coverage [11], could also be used in our approach. We record the input to a method execution as a method-level test when it covers any new coverage element with respect to the chosen coverage criterion. In the second step, we reduce method-level tests with large collection-typed variables using binary reduction. The reduced tests preserve the same coverage achieved by the originally recorded method-level tests. During debugging, a developer will first identify suspicious methods based on his or her understanding of the program. Then, the developer will only need to re-execute the reduced method-level tests recorded for these methods, instead of executing the entire application with the original dataset. Doing so could significantly speed up the debugging process.

We conducted an experimental evaluation of our approach. In our experiments, we selected seven methods from four machine learning algorithms that were implemented in WEKA using Java. The four machine learning algorithms from WEKA and two datasets from UCI dataset repository were selected based on the execution time and size of datasets. Method-level tests were recorded for these seven methods based on three coverage criteria, including edge coverage, edge-pair coverage, and edge-set coverage. (The three coverage criteria are defined

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

in Section II-A.) On average, 4.4 tests were recorded for edge coverage, 5.9 tests for edge-pair coverage, and 18.6 tests for edge-set coverage. While initially, the seven methods were executed from 191 to half a billion times. For some of the recorded method-level tests with large-size inputs, e.g., the previously mentioned *updateStatsForClassifier* method in the *DecisionTable* algorithm, we further reduced the size of the inputs using a binary reduction technique while preserving the same coverage achieved by the original method-level test. For example, the average input size for *updateStatsForClassifier* was reduced to 12.53 MB from 1269.76 GB.

Moreover, test effectiveness was evaluated using PITest (PIT) [16], a commonly used mutation testing tool. Mutation testing seeds faults in a systematic manner to simulate mistakes that developers may make during programming. All 25 available mutant generators were enabled for mutant generation. When combining each set of tests generated for the edge, edge-pair, and edge-set coverage for each method, the mutant killing rate ranges from 57.58% to 91.43%.

We summarize the contributions of our paper as follows:

- We present a new framework for debugging big data applications based on method-level tests. Compared to executing the original dataset at the system level, these method-level tests can be much faster to execute and inspect, which could significantly speed up the debugging process.
- We built a prototype that implements our framework and conducted an experimental evaluation of the framework. The evaluation results suggest that our framework could significantly reduce the time and effort required for debugging big data applications.

The rest of the paper is organized as follows. Section II presents the details of our approach and discusses several implementation challenges. Section III presents the experimental design and analysis of the experimental results. Section IV provides an overview of existing work that is closely related to ours. Section V provides concluding remarks as well as several directions for our future work.

II. Approach

Our approach consists of two major steps, recording method-level tests and reducing the size of the recorded tests. In this section, Section II-A presents our approach to recording method-level tests based on a given coverage criterion. Section II-B presents our approach to reducing the size of a recorded test.

A. Record Test

In a typical scenario, once a failure occurs, a developer identifies several suspicious locations based on his or her understanding of the program. Next, the developer could set up breakpoints in these locations and then start the debugging process with the system-level inputs. The breakpoints allow the developer to inspect the program state during the debugging process. This approach may not be effective for big data applications. This is because when the dataset is large, a breakpoint may be executed for a large number of times before an incorrect program state is found, and each breakpoint has to be inspected manually.

In our approach, the developer first identifies suspicious methods, in a way that is similar to the identification of suspicious locations. Next, our approach runs the program with the original dataset and records, for each suspicious method, a small number of method executions, which we refer to as method-level tests, based on a specific coverage criterion. The method-level tests recorded for a given method achieve the same coverage criterion as the original dataset for the method. The developer can then debug each method with the recorded method executions, instead of a potentially large number of method executions. Since the same coverage criterion is satisfied, there is a high probability that debugging these recorded method-level tests would allow us to detect the fault that may have caused the failure observed at the system level.

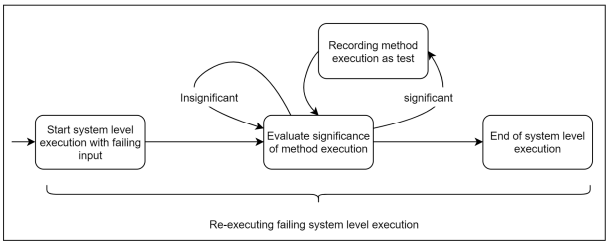


Fig. 1. Recording Process at Runtime

After the developer identifies a list of suspicious methods to be recorded, we instrument these methods to capture the coverage elements that need to be covered for the selected coverage criterion. After instrumentation, our recording process at runtime is shown in Figure 1. While re-executing the failing system-level execution, each method execution of the suspicious methods is evaluated to determine whether it is significant based on the selected coverage criterion. A method execution is considered to be significant if it covers at least one new coverage element. When a method execution is deemed to be significant, its corresponding input for reproducing the method execution is recorded as a method-level test. Otherwise, the execution will continue until it reaches the next significant method execution.

In this paper, we will use edge coverage [11], edge-pair coverage [11], and edge-set coverage, as the coverage criteria based on Control Flow Graph (CFG) to determine if a given method execution is significant. A CFG is a graphical representation of all possible paths that might be traversed by a program at runtime. Thus it captures information about how the control is transferred in a program.

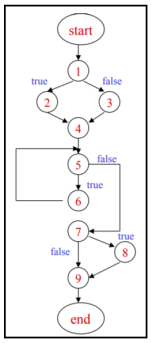
Figure 2 shows an example CFG. In a CFG, each node in the graph represents a basic block, i.e. a sequence of consecutive statements with a single entry and a single exit point[11]. A directed edge [11] represents that the control can flow from one node to another. And a path [11] is a sequence of nodes, where each pair of adjacent nodes is an edge.

We record the method executions as method-level tests when they cover any new coverage elements with respect to

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

the chosen coverage criterion. For edge coverage, each edge covered by a method execution is recorded for the method evaluation. For edge-pair coverage, each edge-pair (reachable path of length up to two) is recorded for the method evaluation. Note that when edge-set coverage is used, a method execution is considered significant if it covers a unique set of edges, i.e., no other method executions exactly cover the same set of edges. Also note that other coverage criteria, e.g., prime-path coverage [11], could also be used in our approach.

```
1.  begin
2.    int x, y, power;
3.    float z;
4.    input (x, y);
5.    if (y < 0)
6.       power = -y;
7.    else
8.       power = y;
9.    z = 1;
10.   while (power != 0) {
11.      z = z * x;
12.      power = power – 1;
13.   }
14.   if (y < 0)
15.      z = 1/z;
16.   output (z);
17. end;
```



Block	Lines	Entry	Exit
1	2, 3, 4, 5	2	5
2	6	6	6
3	8	8	8
4	9	9	9
5	10	10	10
6	11, 12	11	12
7	14	14	14
8	15	15	15
9	16	16	16

Fig. 2. Example of Control Flow Graph

To record method-level tests, three major tasks need to be accomplished, including instrumentation, method execution evaluation, and serialization. We further discuss these tasks in the following subsections.

1) Instrumentation

We use a tool called Atlas [15], which is an Eclipse plugin developed by EnSoft Corp to automatically generate CFGs from the source code of a selected method. Atlas uses each line of code as a basic block. This is different from the classical definition [11] that a basic block consists of a sequence of consecutive statements with a single entry and a single exit point. Figure 2 shows a simple method and its CFG generated using Atlas. We modify the generated CFGs from Atlas by combining blocks that are in a consecutive sequence without inner branches. Doing so reduces the amount of instrumentation and thus the runtime overhead when executing the instrumented code. The red rectangle in Figure 3 marks the lines of code combined to be a basic block as we previously defined.

```
public float power(int x, int y){
    int power;
    float z;
    if (y < 0){
        power = -y;
    }else{
        power = y;
    }
    z = 1;
    while (power != 0){
        z = z * x;
        power = power - 1;
    }
    if(y < 0){
        z = 1 / z;
    }
    return z;
}
```

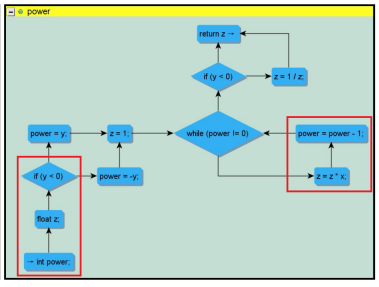


Fig. 3. Example of Modifying Generated Control Flow Graph

Once we have the CFG of a suspicious method, we instrument the method by adding a few lines of code that invokes our recording program. Figure 4 shows an example of how we instrument a sample method. The highlighted statements are extra code added by instrumentation. The code from line 3 to line 10 initializes the recording process. They are inserted at the beginning of a suspicious method. The *ParaArray* array contains the list of input parameters used for a method execution. The *ParaTypeArray* array contains the object types of the input parameters, which are needed to reload the recorded inputs using Java Reflection. When recording a method execution, we record not only the input parameters but also the current object on which the suspicious method was invoked, to store the instance variables accessed during the execution. They are loaded into our system using the "*R.loadInputs(ParaArray, this);*" statement. The statement "*R.enterBlock(#number);*" is added before each basic block to record the index of the basic block when it is executed. The block number *#number* is manually determined based on the previously discussed CFG. Moreover, the statement "*R.endOfProcess();*" is added before each return statement or at the end of a method to notify our program a method execution is completed, and start the method execution evaluation process.

```
1  public float power(int x, int y) {
2      //Insert @ Beginning Of Method
3      RecordMethodExecution R = new RecordMethodExecution();
4      Object[] ParaArray = {x, y};
5      Class[] ParaTypeArray =
6              new Object() {}.getClass().getEnclosingMethod().getParameterTypes();
7      String MethodName = this.getClass().getName() + "/" +
8              new Object() {}.getClass().getEnclosingMethod().getName();
9      R.initializeProcess(MethodName, ParaTypeArray);
10     R.loadInputs(ParaArray, this);
11     //Insert @ Beginning Of Each Block
12     R.enterBlock(0);
13     int power;
14     float z;
15     if (y < 0) {
16         R.enterBlock(1);
17         power = -y;
18     }else {
19         R.enterBlock(2);
20         power = y;
21     }
22     R.enterBlock(3);
23     z = 1;
24     while (power != 0) {
25         R.enterBlock(4);
26         R.enterBlock(5);
27         z = z * x;
28         power = power - 1;
29     }
30     R.enterBlock(4);
31     R.enterBlock(6);
32     if (y < 0) {
33         R.enterBlock(7);
34         z = 1 / z;
35     }
36     R.enterBlock(8);
37     //Insert @ Before Each Return or End of Method
38     R.endOfProcess();
39     return z;
40 }
```

Fig. 4. Example of Instrumentation

Recording basic block indexes with multiple entrances at runtime requires more work than just adding the "*R.enterBlock(#number)*" statement in front of it. As shown in Figure 4, lines 25 to 26 and lines 30 to 31 are the extra codes added for recording the basic block contains line 24. To record the basic block indexes correctly for basic blocks with multiple entrances such as for *while* loop, *for* loop, *else if*, and *switch* statements, etc., we are inserting the "*R.enterBlock(#number);*" statement before its descendants' "*R.enterBlock(#number);*" statement based on the CFG to capture every execution of such blocks. For example, if we only add "*R.enterBlock(#number);*"

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

statement right before the *while* statement shown in Figure 4 at line 24, when the loop comes back to re-evaluate the loop condition at the *while* statement, the repeated execution of this block will not be captured.

2) Method Execution Evaluation

In our implemented framework, we temporarily store the covered edges, edge-pairs, and edge-set for each method execution. We consider a method execution to be significant, and thus record the execution as a method-level test if it covers any edge, edge-pair or edge-set that has not been covered before. Note that we check for uncovered edges first for each method execution. This is because if a method execution covers any edge that has not been covered before, it must cover some new edge-pair(s) and a new edge-set. The time complexity for evaluating each method executions is $O(n^2)$ where n represents the number of coverage elements each method execution has to evaluate. For each method execution, each coverage element of the method execution will be compared to the list of the previously covered elements. If a method execution covers any new coverage element, the method execution will be recorded, and the newly covered elements will be added to the list.

3) Serialization

Once a method execution is determined to be significant, we record the inputs of the method execution using serialization. Serialization can be an expensive process, the built-in serialization support in Java is rather slow when serializing large objects. We used an alternative tool called FST [17] that can be ten times faster [17] to improve the performance of our test recording. In our experiments, FST was able to serialize and deserialize objects correctly. However, there are some reported cases [17] where FST was unable to correctly serialize and deserialize objects that the built-in Java serialization could. In comparison, FST provides better performance, but FST does not provide serialization ability that is as strong as the Java built-in serialization.

While our performance is improved using FST, there are still some situations where we experience significant overhead. To ensure an exact copy of the input objects is created, we perform deep copy on the objects by serializing and deserializing these objects. This is needed because the value of an input object could potentially change during a method execution, especially for void methods that operate on instance variables.

However, most of the stored input objects will not be recorded if the method execution does not cover any new coverage element. Thus, much of the time spent to store the deep copies of objects is unnecessary. These unnecessary time can be huge when a method takes large inputs and/or is executed for a large number of times. The recording overhead can be as high as 7 to 30 times the original system-level execution time for some of the selected methods. In such cases, our solution is recording the method-level tests by executing the entire system twice. In the first execution, we do not store any inputs. Instead, we only record the IDs of significant method executions. In the second execution, we only serialize the selected method executions to store their inputs as method-level tests. Doing so can significantly reduce the runtime overhead in cases where a method takes large inputs or is being executed for a large number of times.

B. Test Reduction

While the recorded method-level tests can be used for debugging, these tests in some case consist of very large inputs. For example, one of the selected methods *cutPointsForSubset,* its recorded method-level tests have the average size of 1.62GB, executing these tests can take a lot of time. And breakpoints in loop statements can be executed for a large number of times. These inputs are large mostly due to the fact that they contain large collections of objects. For the three methods mentioned above, they all have *Instances* typed (Implements *Collection*) variables that contain instances from the original dataset for processing. Some of the recorded data could potentially be reduced while still reproducing the method execution and preserving the coverage elements. The reduction can further reduce the time for executing the tests, and the debugging efforts required from developers.

Our binary reduction technique is inspired by the commonly used binary search technique. For each recorded method-level test, we divide its collection typed input variables into halves. Next, we take each half and other non-collection typed inputs and re-execute them with the suspicious method. We then check whether a half can preserve the originally covered coverage elements. If one of the halves does preserve all the coverage elements, we will continue dividing it into halves and check for the coverage elements repeatedly, until the minimal subset of the collection variables that can preserve the coverage elements are identified. Note that when preserving the coverage during reduction, we are preserving the exact covered elements of edge coverage, edge-pair coverage, and edge-set coverage.

III. Experiments

We implemented the initial working prototype of our framework in Java. Some Manual efforts are required from developers to instrument the source code of suspicious methods. After instrumentation, the recording process has been automated. The reduction approach requires developers to manually identify the large collection typed input variables. The re-execution of the recorded and reduced method-level tests has been automated for debugging. We also conducted mutation testing to evaluate the fault detection effectiveness of our recorded and reduced method-level test. The currently implemented coverage criteria are the edge, edge-pair, and edge-set coverage.

In the following, we discuss how we conducted our experiments and present the experiment results. In Section III-A, we discuss how we selected datasets, applications, and methods to be used for our experiments. Section III-B presents the statistics of the recorded method-level tests. Section III-C presents the statistics of the reduced method-level tests. Section III-D presents how we conducted a mutation testing experiment and the results of our mutation testing for both the recorded tests, and the reduced tests. And finally, Section III-E presents the performance analysis of our framework. All the

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

source code, recorded method-level tests, reduced method-level tests and mutation reports are publicly available at https://www.dropbox.com/sh/3k4kjwqjpa9i2qv/AAAkeYYNaQOVfT9WGe4OUp_Pa?dl=0 for review. The machine we used for our experiment is a workstation with two Xeon E5-2630V3 8 core CPUs @ 2.40GHz, 64GB DDR4 2133 MT/s memory, and a Samsung 850 EVO 500GB SSD.

A. Subjects

We design our experiments to reflect real-world situations for evaluating the effectiveness of our framework. First, we randomly selected ten algorithms that are implemented in the WEKA tool. WEKA is one of the most widely used tools for data mining by practitioners. Next, we selected one collection of dataset with the largest number of instances (accessed on 08/18/2018) from the UCI Machine Learning Repository that consists of 440 real-world collected datasets as a start. The selected collection of datasets, *Heterogeneity Activity Recognition* (HAR), contains four datasets for four different types of devices with a total of 43,930,257 instances and 16 attributes. The HAR collection includes several data types, including multivariate, time-series and real numbers. The datasets can be used for both classification and clustering. Among the four datasets, the largest dataset, *Phones_gyroscope*, is used to execute the ten algorithms.

Phones_gyroscope dataset has the size of 1.37GB, it is too large for two of our selected algorithms *EM* and *LibSVM* to finish their execution within a day. The execution time is too long for our experimentation purpose due to our limited time and resources. For these two algorithms, we reduced the size of the *Phones_gyroscope* dataset by dividing the dataset in half and continue to divide in half until the execution time for *EM* and *LibSVM* are reduced to be near an hour. The reduced *Phones_gyroscope* dataset for *EM* and *LibSVM* now has the size of 3.3 MB. *EM* will now take 5352 seconds (1.49 Hours) to execute and 4491 seconds (1.25 Hours) for *LibSVM*.

TABLE I. SELECTED METHOD INFORMATION

Method	Algorithm	# of Covered Lines of Code	# of Total Lines of Code	# of Execution Count
buildClusterer	EM	115	165	1,910
cutPointsForSubset	DecisionTable	62	64	29,564
EM_Init	EM	47	53	191
handleNumericAttribute	J48	51	53	28,314
select_working_set	LibSVM	50	52	417,989
selectModel	J48	50	58	12,391
updateStatsForClassifier	DecisionTable	46	66	557,305,280

After two datasets (original *Phone_gyroscope* dataset and the reduced dataset) and ten algorithms' implementations (*Apriori, DecisionTable, EM, HierarchicalClusterer, J48, LibSVM, LinearRegression, MakeDensityBasedClusterer, RandomTree, SimpleKMeans*) have been selected. We select methods with a larger number of executed statements, and a larger number of executions for our experiments. This is because longer methods and methods that have been executed for a larger number of times often require more effort to debug. A total of seven methods are selected. The selected methods and their information are shown in Table I. These methods are then instrumented as previously described in Section II.

B. Recorded Method-Level Tests

For our experiments, we have recorded method-level tests for all of the seven selected methods for preserving edge coverage, edge-pair coverage, and edge-set coverage of the original system-level execution. Some important information about the recorded method-level tests is shown in Table II. Note that the statement coverage column in Table II is for all three types of recorded tests, as well as the original failing system-level execution. This is because edge coverage subsumes statement coverage, once all edges are preserved, all the statement coverage will be preserved as well, and edge-pair coverage and edge-set coverage both subsume edge coverage.

Based on the results shown in Table II, we can see that only a small number of method-level tests are sufficient for preserving coverage for a suspicious method. Empirical studies show that there exists a high correlation between code coverage and fault detection effectiveness. The actual fault detection ability of our recorded method-level tests will be further evaluated using mutation testing in Section III-D. Thus, when failures occur on a system level, it is likely that executing the method-level tests for the suspicious methods would trigger the failure observed during the execution with the original dataset. Thus, the use of method-level tests could potentially save developers a lot of time and efforts.

C. Reduced Method-Level Tests

As shown in Table III, while some of the tests have a reasonable size, three methods, *cutPointsForSubset*, *selectModel* and *updateStatsForClassifier* have significantly large inputs for their recorded method-level tests. While debugging with these tests is easier than debugging with the original dataset at the system level, loading and debugging these tests could still take a lot of time. We further reduce the size of these tests using our binary reduction approach as discussed in Section II. In Table III, we compare the differences between the recorded method-level tests before and after they were reduced.

For size reduction, our binary reduction technique was able to reduce the input size of tests for five out of seven methods. Our result shows that the reduction amount is often above 95%. Most of the method-level tests can be reduced significantly while still preserving our selected coverage elements. The coverage element refers to the edges, edge-pairs, and edge-set covered by each recorded method-level test. While one of the tests for *selectModel* can be reduced to 1.7 KB from 1.63 GB, some tests still have a fair amount of input data remaining, such as the reduction from 1.63GB to 37.22 MB for one of the

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

TABLE II. RECORDED METHOD EXECUTION INFORMATION

Method	Edge Coverage		Edge-Pair Coverage		Edge-Set Coverage		Total # of Recorded Tests	# of Original Execution Count	Statement Coverage
	# of Tests	# of Covered Edges	# of Tests	# of Covered Edge-Pairs	# of Tests	# of Covered Edge-Sets			
buildClusterer	3	83	4	181	3	3	4	1,910	69.70%
cutPointsForSubset	8	30	9	64	17	17	18	29,564	96.88%
EM_Init	1	24	3	55	1	1	3	191	88.68%
handleNumericAttribute	4	32	5	70	33	33	33	28,314	96.23%
select_working_set	7	41	11	111	61	61	63	417,989	96.15%
selectModel	5	35	5	74	6	6	6	12,391	86.21%
updateStatsForClassifier	3	26	5	59	9	9	11	557,305,280	69.70%

tests of *cutPointsForSubset*. Furthermore, we were unable to reduce any test inputs for two methods, *buildClusterer* and *EM_Init*. We further investigated this by looking into how the variables of collection type are accessed and used. We noticed mainly three different scenarios that may have contributed to our results.

The first scenario is when a collection variable is partially used as inputs. When the partially accessed instances are in a consecutive sequence in the collection variable, or when only one instance is accessed, our binary reduction technique will reduce such collection variable to its minimal subset. However, if the accessed instances are spread across the collection variable, our binary reduction will not be able to identify only the accessed instances. Hence, the reduction may not be minimal, many unnecessary data based on the coverage elements may remain.

```
744       if (m_executionSlots <= 1
745         || instances.numInstances() < 2 * m_executionSlots) {
746         for (i = 0; i < instances.numInstances(); i++) {
747           Instance toCluster = instances.instance(i);
748           int newC =
749             clusterProcessedInstance(
750               toCluster,
751               false,
752               true,
753               m_speedUpDistanceCompWithCanopies ? m_dataPointCanopyAssignments
754                 .get(i) : null);
755           if (newC != clusterAssignments[i]) {
756             converged = false;
757           }
758           clusterAssignments[i] = newC;
759         }
760       } else {
761         converged = launchAssignToClusters(instances, clusterAssignments);
762       }
```

Fig. 5. Collection Variable Used at Branching Condition

The second scenario is when the collection variable is accessed in branching statements, e.g. for the tests recorded for *buildClusterer* and *EM_Init*. The collection variables identified for these two methods were used at a few branching statements and passed to other methods that return value to the execution as well. In this situation, maintaining the exact coverage elements can be difficult to achieve for our binary reduction technique. As an example, part of the code of *buildClusterer* is shown in Figure 5. The *instances* variable was used at an *if* statement and in the conditions of a *for* loop. Reducing the *instance* variable using our binary reduction approach will compromise the originally covered coverage elements (edges, edge-pairs, edge-set) of the method-level tests recorded for the *buildClusterer* method.

The third scenario is when the collection variable is not accessed at all. In our implementation, to reduce manual efforts required for instrumentation and reproduce method executions precisely, we automatically record both the parameters passed to the method and the object where the method was invoked from, ensuring all possible inputs are recorded. However, not all recorded information is used as inputs, such as for some instance variables of the object where the method was invoked from. In this situation, our binary reduction technique may be able to reduce unnecessary collection variables to empty, while still preserving the coverage elements.

The first and second scenario can potentially use delta debugging [8] or preserving superset of the coverage elements to further the reduction. However, delta debugging could significantly increase the reduction overhead, and preserving superset of the coverage elements may lose or introduce some coverage elements that could potentially have a large impact on the reduced method-level test. For the third scenario, we can implement systematic static analysis in the future to help our framework identify and record only the necessary inputs for reproducing method executions.

For execution time reduction, many of the recorded set of method-level tests are now taking seconds instead of minutes after the binary reduction. When debugging with these reduced tests, not only the tests will be short and easier to debug, the execution time is also easy to manage.

D. Mutation Testing

For mutation testing, we used PITest (PIT) [16], a mutation testing tool for Java, to evaluate the fault detection effectiveness of our recorded method-level tests. In PIT, different types of faults (or mutants) are automatically seeded

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

TABLE III. TEST REDUCTION RESULTS

Method	Total # of Tests	# of Large Collection Variables	Average Input Size (MB)		Total Input Size (MB)		Maximum Input Size (MB)		Minimum Input Size (MB)		Test Execution Time (Seconds)	
			Recorded	*Reduced*	*Recorded*	*Reduced*	*Recorded*	*Reduced*	*Recorded*	*Reduced*	*Recorded*	*Reduced*
buildClusterer	4	1	3.18	3.18	12.75	12.75	3.23	3.23	3.16	3.16	5 s	5 s
cutPointsForSubset	18	2	1628.16	9.77	29306.81	176.25	1628.16	37.22	1628.16	0.001	1836 s	5 s
EM_Init	3	1	4.71	4.71	14.12	14.12	4.34	4.34	5.18	5.18	5 s	5 s
handleNumericAttribute	33	1	54.00	0.74	1781.76	24.42	1300.48	1.75	0.0012	0.001	155 s	2 s
select_working_set	63	2	39.50	0.08	2488.32	5.04	41.9	0.29	1.83	0.002	176 s	1 s
selectModel	6	2	1392.64	0.02	8357.04	0.11	1628.16	0.01	1320.96	0.001	682 s	1 s
updateStatsForClassifier	11	1	1269.76	12.53	13967.34	138.06	1269.76	44.86	1269.76	0.51	875 s	3 s

into the source code. Each mutation (a mutated version of source code) simulates a single fault and is executed against the unit tests that developers provide.

Mutation testing requires the provided unit tests to be passing tests. This is because only when the mutant's output differs from the expected output, a mutant is said to be killed. In our experiments, when a method-level test is executed, we record the outputs as the expected output for mutation testing purpose. The output for each test contains not only the returned object if there is one, but also the object where the method was invoked from and the input parameters of the method. This is because the values of these parameters and the object where the method was invoked from could change and should be considered as part of the output.

PIT provides a total of 25 different mutators to mutate different type of code. When conducting mutation testing, we have enabled all 25 mutators in PIT for generating mutants in our selected methods. PIT also provides an option to set a timeout factor for executing each test against each mutant. The default is 1.25 times the original test execution time. We increased the timeout factor to 10 times the original execution time, as an effort to avoid false positives killing of mutants. This is because a timed-out mutant is also considered as a killed mutant. We have also increased the Java heap size to 60GB and stack size to 128MB using JVM configuration in PIT, to avoid false positive killing of memory error mutants.

Table IV shows the mutation testing result of our recorded and reduced method-level tests. Note that PIT currently does not support the mutant generation of only covered statements. Because the mutation generation of PIT is done statically, it will generate mutants for all the statements of a selected method, instead of only the reachable ones. In other words, if a mutant is located at a statement that was not covered by any of the tests, the mutant will not be exercised, and thus is impossible to be killed. Such mutants will not be considered in our experiments. This is because if a mutant is not exercised by our recorded tests, it is not exercised by the original system-level execution. The total number of mutants generated for each selected method in Table IV are calculated manually which consist of only exercised mutants by our tests. This is done by removing mutants that are labeled as *NO_COVERAGE* in the mutation testing report generated using PIT, such as shown in Figure 6.

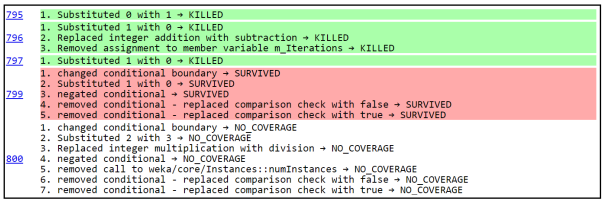


Fig. 6. Sample Mutation Testing Report

For recorded method-level tests without reduction shown in Table IV, we can see that most of the recorded tests for different methods and coverage criteria have a high mutant killing rate. Even without comparing to the original system-level execution, a small number of tests show high effectiveness in detecting potential faults that could occur in the selected methods. For four out of seven selected methods, recorded tests achieve over 80% of mutant killing rate for all the selected coverage criteria. The average mutant killing rate across seven methods are around 80% for all four different sets of tests that achieve edge coverage, edge-pair coverage, edge-set coverage, and these three combined. By only using edge coverage, the recorded method-level tests can achieve reasonably high mutant killing rate. With edge-pair and edge-set coverage, the mutant killing rate is further improved slightly in some cases. This indicates the method-level tests generated using our framework can effectively help developers to debug and find faults they are looking for, while significantly reducing the time and efforts required from developers for debugging.

For reduced method-level tests, their mutant killing rates are nearly the same as their original recorded tests. With differences no larger than 5% of their original killing rate. We even see some cases with increased mutant killing rate, such as for the edge-pair coverage of method "*cutPointsForSubset*". While coverage elements of our

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

TABLE IV. MUTATION TESTING RESULTS

Method	# of Mutants Generated for Covered Code	Statement Coverage	Edge Coverage Mutant Killing Rate		Edge-Pair Coverage Mutant Killing Rate		Edge-Set Coverage Mutant Killing Rate		Combined Recorded Tests Mutant Killing Rate	
			Recorded	*Reduced*	*Recorded*	*Reduced*	*Recorded*	*Reduced*	*Recorded*	*Reduced*
buildClusterer	269	69.70%	79.18%	79.18%	79.18%	79.18%	79.18%	79.18%	79.18%	79.18%
cutPointsForSubset	164	96.88%	81.71%	81.1%	81.71%	82.32%	85.98%	85.98%	85.98%	85.98%
EM_Init	102	88.68%	87.25%	87.25%	87.25%	87.25%	87.25%	87.25%	87.25%	87.25%
handleNumericAttribute	140	96.23%	89.29%	88.57%	90.71%	90.71%	91.43%	91.43%	91.43%	91.43%
select_working_set	128	96.15%	71.88%	75%	73.44%	75%	75%	78.91%	75%	78.91%
selectModel	132	86.21%	57.58%	57.58%	57.58%	57.58%	62.88%	62.88%	62.88%	62.88%
updateStatsForClassifier	122	69.70%	86.07%	81.15%	86.07%	84.43%	86.89%	86.07%	86.89%	86.89%
Average			79.00%	78.55%	79.42%	79.49%	81.23%	81.67%	81.23%	81.79%

specifically selected coverage criteria are maintained, other elements from other coverage criteria could become lost, or may be newly introduced after our binary reduction, such as combinations of the different branches being executed. The mutation testing results of the reduced tests show that even after the input sizes are significantly reduced, the coverage elements and also the fault detection effectiveness are still preserved. Our binary reduction technique on method-level tests can further help developers to reduce efforts for debugging while maintaining the debugging effectiveness of the method-level tests.

TABLE V. SYSTEM-LEVEL MUTATION TESTING

Method	Algorithm	# of Mutants Killed by System-Level Execution	# of Propagatable Mutants Killed by Combined Method-Level Tests
select_working_set	LibSVM	58	51
selectModel	J48	61	56

We also investigated the two methods *select_working_set* and *selectModel* with the lowest mutant killing rate by comparing their results to the mutation testing results of their system-level execution. We have planned on comparing all recorded method-level tests' mutation testing results with their corresponding system-level execution. However, while mutation testing is a very effective method to evaluate the quality of tests, mutation testing is a rather expensive method to use. In this paper, we only have two system-level mutation testing results for *select_working_set* and *selectModel*. Moreover, their system-level mutation tests both took over one week to complete. Note that some mutants that can be killed with method-level tests are not propagatable on the system level, i.e., a mutant may cause a method execution producing incorrect output, but such incorrect output on the method level did not cause an incorrect system-level output. We considered the option of recording all method executions of a method during its system-level execution. However, it is impractical, because of our selected methods have been executed with a large number of times, and many of them have large inputs as well. For comparing mutation testing results between method-level tests and system-level execution, we will only be considering the propagatable mutants for the method-level tests.

The system-level mutation testing results for *select_working_set* and *selectModel* are shown in Table V. For *LibSVM*, the system-level execution was able to kill 58 mutants, the combined method-level test of *select_working_set* was able to kill 51 out 58 propagatable mutants with a propagatable mutant killing rate of 87.93%. For *J48*, the system-level execution was able to kill 61 mutants, the combined method-level tests of *selectModel* were able to kill 56 out of 61 propagatable mutants with a propagatable mutant killing rate of 91.80%. The further investigation shows the reason why method-level tests recorded for *select_working_set* and *selectModel* have a lower mutant killing rate. It is likely because their original system-level execution has a lower mutant killing rate.

After investigating the un-killed propagatable mutants in the recorded method-level tests, we discovered three un-killed propagatable mutants from *select_working_set* and one from *selectModel* were mutations related to modifying boundary conditions. This means by adding more coverage criteria related to boundary conditions, a higher mutant killing rate can be achieved for the method-level test. With a few basic coverage criteria implemented for our framework, method-level tests produced by our framework can be very effective in detecting faults during debugging.

E. Performance Evaluation

We evaluate the performance of our implementation by investigating the original system-level execution time, the

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

time taken to evaluate and record the method-level tests, time taken to reduce tests, and the time taken to execute the recorded method-level tests. The results are shown in Table VI. Recall that in the experiments for mutation testing, both inputs and outputs of the selected method executions are recorded. However, the results shown in Table VI are only for recording the inputs and executing the recorded method-level tests with only inputs without comparing their outputs. This is because, in real-world use of our framework, outputs of the method executions do not need to be recorded.

As previously mentioned in Section II, we have two solutions for recording selected method executions. One approach is to serialize and temporarily store the inputs for each method execution and record the inputs locally when a method execution is determined to be significant. This method requires executing the entire system only once. However, in cases where a method has large inputs or is executed for a large number of times, this approach may have a significant performance issue due to all the unnecessary serialization. The other approach is to execute the entire system twice. In the first execution, we evaluate each method execution and store the execution IDs of the method executions. An execution ID is the index of a method execution based on the order of each method executions that happened during the system-level execution. In the second system-level execution, we only serialize and record the inputs of the selected method executions based on their execution IDs. The numbers marked with "*" indicates that the method-level tests were recorded using the second recording approach as shown in Table VI. The execution time is computed by subtracting the execution end time by the execution start time that was created using the Java *System.CurrentTimeMillis()* function.

TABLE VI. PERFORMANCE EVALUATION RESULTS

Method	Original Execution Time	Total Test Recording Time	Total Test Execution Time		Total Test Reduction Time
			Recorded	*Reduced*	
buildClusterer	5352 s	6303 s	5 s	5 s	27 s
cutPointsForSubset	9559 s	*21615 s	1836 s	5 s	7558 s
EM_Init	5356 s	5361 s	5 s	5 s	22 s
handleNumericAttribute	6357 s	*14624 s	155 s	2 s	1965 s
select_working_set	4491 s	*11531 s	176 s	1 s	2763 s
selectModel	6357 s	*14212 s	682 s	1 s	3122 s
updateStatsForClassifier	9559 s	*30513 s	875 s	3 s	4088 s

In Table VI, we see that recording method-level tests using our framework can take up to three times of the initial system execution. Additional test reduction time could take as much as two hours based on the size of the inputs (Our binary reduction utilizes serialization for deep copy as well). The reduced tests can be executed for many times during the debugging, the reduction time is a one-time investment, we believe the time is manageable for developers. Moreover, our approach is automated, allowing developers to work on other tasks while running our approach. For executing the recorded method-level tests, we see that it usually takes much less time than executing the entire system, especially for the reduced tests, the execution time can range from as little as one second to five seconds. Overall, we believe that recording and reducing method-level tests using our framework will help developers save a lot of time and efforts in debugging big data applications.

IV. RELATED WORK

We first review previous work related to generating tests for big data applications. Csallner et al. proposed an approach that uses dynamic symbolic execution to automatically generate tests for general *MapReduce* programs [1]. Morán et al. proposed *MRFlow,* a testing technique tailored to test *MapReduce* programs [5]. *MRFlow* uses data flow test criteria and oriented to transformations analysis between the input and the output in order to detect defects in *MapReduce* programs. Morán et al. also proposed a technique to generate different infrastructure configurations for a given *MapReduce* program that can be used to reveal functional faults [4]. They also proposed an automatic test framework that can detect functional faults automatically [3]. Chandrasekaran et al. proposed an approach to generate test input data using combinatorial testing for testing big data applications [6]. Previous work reported in [1, 2, 3, 4, 5] focuses on generating tests that help to identify functional faults, i.e., faults that will cause the program to generate unexpected outputs. In contrast, our work focuses on reducing debugging efforts for big data applications. Our tests are recorded in an effort to reproduce failures using a small number of method-level tests.

Second, some work has been reported on debugging big data applications. Gulzar et al. developed a tool, BigDebug, that simulates breakpoints to enable a developer to inspect a program without actually pausing the entire computation [7]. To help a user inspect millions of records passing through a data-parallel pipeline, BigDebug provides *guarded watchpoints*, which dynamically retrieve only those records that match a user-defined guard predicate. Chandrasekaran et al. proposed a technique that uses different annotators to debug the tracking data independently and their debugging results were collected for joint correction propagation for later analysis [9]. Our work is similar to Gulzar [7] and Li [9] in terms of only focusing on a subcomponent of the system. However, our work focuses on recording significant method-level executions to be replayed for debugging suspicious methods. Gulzar [7] and Li [9] focuses on tracking the changes made to certain objects using data flow analysis approach.

Third, our work is also related to existing work that records program information and uses the information to generate unit tests. Pasternak et al. proposed a technique that records interactions that take place during the execution of Java programs and uses these interactions to construct unit tests automatically using GenUTest [10]. Orso et al. proposed

Feng, Huadong; Chandrasekaran, Jagan; Lei, Yu; Kacker, Raghu; Kuhn, David.
"A Method-Level Test Generation Framework for Debugging Big Data Applications."
Paper presented at 2018 IEEE International Conference on Big Data, Seattle, WA, United States. December 10, 2018 - December 13, 2018.

a technique and conducted a feasibility study using SCARPE, a prototype tool, for selective capture and replay of program executions [6]. Similar to our work presented in this paper, Orso's technique [6] can be used to automatically generate unit tests based on the recorded information for testing purpose. Our work is similar to Pasternak [10] and Orso [6] in terms of recording method-level tests based on the system-level execution. However, our work focuses on recording unit tests for debugging one or more failures that have been observed instead of generating tests for triggering failures that have not been observed yet. Furthermore, our work also does not require complex instrumentation techniques on the target's bytecode [6]. Instead, we only employ simple instrumentation that keeps track of code coverage.

Finally, we review work related to reducing input size for the debugging purpose. Zeller et al. proposed Delta Debugging [8] technique to isolate failure-inducing inputs on the system level to reduce work required for debugging. Clause [14] et al. presented a technique based on dynamic tainting for automatically identifying subsets of a program's inputs that are relevant to a failure. These techniques reduce the debugging effort at the system level, in terms that the reduced datasets need to be executed at the system level. This is in contrast with our work that reduces the debugging effort at the method level.

V. Conclusion & Future Work

In this paper, we presented a framework to provide developers with method-level tests that were recorded from a failed system-level execution with the original dataset. These method-level tests preserve a given coverage criterion, e.g. edge, edge-pair, and edge-set coverage, and thus are likely to reproduce the failure observed at the system level. The binary reduction is used to further reduce method-level tests with large input. The set of method-level tests that are provided by our approach could help developers to effectively debug suspicious methods against properties of the original input dataset, and significantly reduce time and effort required for debugging big data applications.

There are two major directions for future work. First, we plan to conduct more experimental evaluation of our approach using more big data applications, datasets, and coverage criteria. Second, we plan to further automate our approach. In particular, we will develop techniques that can fully automate the instrumentation process. Our current approach still needs manual effort in modifying CFG generated by Atlas, inserting code for instrumentation, and identifying collection typed variables for reduction. It is our plan to make the tool publicly available.

VI. Acknowledgment

This work is supported by a research grant (70NANB15H199) from Information Technology Lab of National Institute of Standards and Technology (NIST).

Disclaimer: Certain software products are identified in this document. Such identification does not imply recommendation by the NIST, nor does it imply that the products identified are necessarily the best available for the purpose.

References

[1] Csallner, C., Fegaras, L., & Li, C. (2011, September). New ideas track: testing mapreduce-style programs. In *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering* (pp. 504-507). ACM.

[2] Jaganmohan Chandrasekaran, Huadong Feng, Yu Lei, Richard Kuhn, Raghu Kacker, "Applying combinatorial testing to data mining algorithms", *Software Testing Verification and Validation Workshops (ICSTW) 2017 IEEE Fourth International Conference on-6th International Workshop on Combinatorial Testing (IWCT),* 2017.

[3] Morán, J., Bertolino, A., de la Riva, C., & Tuya, J. (2017, July). Towards Ex Vivo Testing of MapReduce Applications. In *Software Quality, Reliability and Security (QRS), 2017 IEEE International Conference on* (pp. 73-80). IEEE.

[4] Morán, J., Rivas, B., De La Riva, C., Tuya, J., Caballero, I., & Serrano, M. (2016, August). Infrastructure-aware functional testing of mapreduce programs. In *Future Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference on* (pp. 171-176). IEEE.

[5] Morán, J., Riva, C. D. L., & Tuya, J. (2015, August). Testing data transformations in MapReduce programs. In *Proceedings of the 6th International Workshop on Automating Test Design, Selection and Evaluation* (pp. 20-25). ACM.

[6] Orso, A., & Kennedy, B. (2005, May). Selective capture and replay of program executions. In *ACM SIGSOFT Software Engineering Notes* (Vol. 30, No. 4, pp. 1-7). ACM.

[7] Muhammad Ali Gulzar, Matteo Interlandi, Seunghyun Yoo, Sai Deep Tetali, Tyson Condie, Todd Millstein, Miryung Kim. BigDebug: Debugging Primitives for Interactive Big Data Processing in Spark. Proceeding *ICSE '16 Proceedings of the 38th International Conference on Software Engineering*, Pages 784-795

[8] A. Zeller and R. Hildebrandt, "Simplifying and Isolating Failure-Inducing Input", IEEE Transactions on Software Engineering28(2), February 2002, pp. 183-200.

[9] Mingzhong Li, Zhaozheng Yin. Debugging Object Tracking by a Recommender System with Correction Propagation. In IEEE Transactions on Big Data (Volume: 3, Issue: 4, Dec. 1 2017)

[10] Pasternak, B., Tyszberowicz, S., & Yehudai, A. (2009). GenUTest: a unit test and mock aspect generation tool. *International journal on software tools for technology transfer*, *11*(4), 273.

[11] N. Li, U. Praphamontripong, and J. Offutt, "An experimental comparison of four unit test criteria: Mutation, edge-pair, all-uses and prime path coverage," in Second International Conference on Software Testing Verification and Validation, ICST 2009, Denver, Colorado, USA, April 1-4, 2009, Workshops Proceedings, 2009, pp. 220–229.

[12] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[13] Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[14] J. Clause and A. Orso. Penumbra: Automatically identifying failure relevant inputs using dynamic tainting. In ISSTA, pages 249–260, 2009.

[15] "Atlas Platform, EnSoft Corp." http://www.ensoftcorp.com.

[16] "PITest." http://pitest.org/.

[17] "FST, fast-serialization." https://github.com/RuedigerMoeller/fast-serialization.

Constant-Round Group Key Exchange from the Ring-LWE Assumption

Daniel Apon[1], Dana Dachman-Soled[2], Huijing Gong[2], and Jonathan Katz[2]

[1] National Institute of Standards and Technology, USA
daniel.apon@nist.gov
[2] University of Maryland, College Park, USA
danadach@ece.umd.edu, {gong, jkatz}@cs.umd.edu

Abstract. Group key-exchange protocols allow a set of N parties to agree on a shared, secret key by communicating over a public network. A number of solutions to this problem have been proposed over the years, mostly based on variants of Diffie-Hellman (two-party) key exchange. To the best of our knowledge, however, there has been almost no work looking at candidate *post-quantum* group key-exchange protocols.

Here, we propose a constant-round protocol for unauthenticated group key exchange (i.e., with security against a passive eavesdropper) based on the hardness of the Ring Learning With Errors (Ring-LWE) problem. By applying the Katz-Yung compiler using any post-quantum signature scheme, we obtain a (scalable) protocol for *authenticated* group key exchange with post-quantum security. Our protocol is constructed by generalizing the Burmester-Desmedt protocol to the Ring-LWE setting, which requires addressing several technical challenges.

Keywords: Ring learning with errors, Post-quantum cryptography, Group key exchange

1 Introduction

Protocols for (authenticated) key exchange are among the most fundamental and widely used cryptographic primitives. They allow parties communicating over an insecure public network to establish a common secret key, called a *session key*, permitting the subsequent use of symmetric-key cryptography for encryption and authentication of sensitive data. They can be used to instantiate so-called "secure channels" upon which higher-level cryptographic protocols often depend.

Most work on key exchange, beginning with the classical paper of Diffie and Hellman, has focused on two-party key exchange. However, many works have also explored extensions to the *group* setting [21, 29, 15, 30, 5, 6, 25, 14, 12, 13, 11, 17, 22, 16, 8, 2, 1, 24, 9, 31] in which N parties wish to agree on a common session key that they can each then use for encrypted/authenticated communication with the rest of the group.

The recent effort by the National Institute of Standards and Technology (NIST) to evaluate and standardize one or more quantum-resistant public-key

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

cryptosystems is entirely focused on digital signatures and two-party key encapsulation/key exchange,[1] and there has been an extensive amount of research over the past decade focused on designing such schemes. In contrast, we are aware of almost no[2] work on *group* key-exchange (GKE) protocols with post-quantum security beyond the observation that a post-quantum group key-exchange protocol can be constructed from any post-quantum two-party protocol by having a designated group manager run independent two-party protocols with the $N-1$ other parties, and then send a session key of its choice to the other parties encrypted/authenticated using each of the resulting keys. Such a solution is often considered unacceptable since it is highly asymmetric, requires additional coordination, is not contributory, and puts a heavy load on a single party who becomes a central point of failure.

1.1 Our Contributions

In this work, we propose a constant-round group key-exchange protocol based on the hardness of the Ring-LWE problem [27], and hence with (plausible) post-quantum security. We focus on constructing an *unauthenticated* protocol—i.e., one secure against a passive eavesdropper—since known techniques such as the Katz-Yung compiler [24] can then be applied to obtain an *authenticated* protocol secure against an active attacker.

The starting point for our work is the two-round group key-exchange protocol by Burmester and Desmedt [15, 16, 24], which is based on the decisional Diffie-Hellman assumption. Assume a group $\mathbb{G}$ of prime order q and a generator $g \in \mathbb{G}$ are fixed and public. The Burmester-Desmedt protocol run by parties $P_0, \ldots, P_{N-1}$ then works as follows:

1. In the first round, each party P_i chooses uniform $r_i \in \mathbb{Z}_q$ and broadcasts $z_i = g^{r_i}$ to all other parties.
2. In the second round, each party P_i broadcasts $X_i = (z_{i+1}/z_{i-i})^{r_i}$ (where the parties' indices are taken modulo N).

Each party P_i can then compute its session key sk_i as

$$\mathsf{sk}_i = (z_{i-1})^{N r_i} \cdot X_i^{N-1} \cdot X_{i+1}^{N-2} \cdots X_{i+N-2}.$$

One can check that all the keys are equal to the same value $g^{r_0 r_1 + \cdots + r_{N-1} r_0}$.

In attempting to adapt their protocol to the Ring-LWE setting, we could fix a ring R_q and a uniform element $a \in R_q$. Then:

1. In the first round, each party P_i chooses "small" secret value $s_i \in R_q$ and "small" noise term $e_i \in R_q$ (with the exact distribution being unimportant in the present discussion), and broadcasts $z_i = a s_i + e_i$ to the other parties.

[1] Note that CPA-secure key encapsulation is equivalent to two-round key-exchange (with passive security).

[2] The protocol of Ding et al.[19] has no security proof; the work of Boneh et al.[10] shows a framework for constructing a group key-exchange protocol with plausible post-quantum security but without a concrete instantiation.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

2. In the second round, each party P_i chooses a second "small" noise term $e'_i \in R_q$ and broadcasts $X_i = (z_{i+1} - z_{i-i}) \cdot s_i + e'_i$.

Each party can then compute a session key b_i as

$$b_i = N \cdot s_i \cdot z_{i-1} + (N-1) \cdot X_i + (N-2) \cdot X_{i+1} + \cdots + X_{i+N-2}.$$

The problem, of course, is that (due to the noise terms) these session keys computed by the parties will *not* be equal. They will, however, be "close" to each other if the $\{s_i, e_i, e'_i\}$ are all sufficiently small, so we can add an additional reconciliation step to ensure that all parties agree on a common key k.

This gives a protocol that is correct, but proving security (even for a passive eavesdropper) is more difficult than in the case of the Burmester-Desmedt protocol. Here we informally outline the main difficulties and how we address them. First, we note that trying to prove security by direct analogy to the proof of security for the Burmester-Desmedt protocol (cf. [24]) fails; in the latter case, it is possible to use the fact that, for example,

$$(z_2/z_0)^{r_1} = z_1^{r_2 - r_0},$$

whereas in our setting the analogous relation does not hold. In general, the natural proof strategy here is to switch all the $\{z_i\}$ values to uniform elements of R_q, and similarly to switch the $\{X_i\}$ values to uniform subject to the constraint that their sum is approximately 0 (i.e., subject to the constraint that $\sum_i X_i \approx 0$). Unfortunately this cannot be done by simply invoking the Ring-LWE assumption $O(N)$ times; in particular, the first time we try to invoke the assumption, say on the pair $(z_1 = as_1 + e_1,\ X_1 = (z_2 - z_0) \cdot s_1 + e'_1)$, we need $z_2 - z_0$ to be uniform—which, in contrast to the analogous requirement in the Burmester-Desmedt protocol (for the value z_2/z_0), is not the case here. Thus, we must somehow break the circularity in the mutual dependence of the $\{z_i, X_i\}$ values.

Toward this end, let us look more carefully at the distribution of $\sum_i X_i$. We may write

$$\textstyle\sum_i X_i = \sum_i (e_{i+1}s_i - e_{i-1}s_i) + \sum_i e'_i.$$

Consider now changing the way X_0 is chosen: that is, instead of choosing $X_0 = (z_1 - z_{N-1})s_0 + e'_0$ as in the protocol, we instead set $X_0 = -\sum_{i=1}^{N-1} X_i + e'_0$ (where e'_0 is from the same distribution as before). Intuitively, as long as the standard deviation of e'_0 is large enough, these two distributions of X_0 should be "close" (as they both satisfy $\sum_i X_i \approx 0$). This, in particular, means that we need the distribution of e'_0 to be different from the distribution of the $\{e'_i\}_{i>0}$, as the standard deviation of the former needs to be larger than the latter.

We can indeed show that when we choose e'_0 from an appropriate distribution then the Rényi divergence between the two distributions of X_0, above, is bounded by a polynomial. With this switch in the distribution of X_0, we have broken the circularity and can now use the Ring-LWE assumption to switch the distribution of z_0 to uniform, followed by the remaining $\{z_i, X_i\}$ values.

Unfortunately, bounded Rényi divergence does not imply statistical closeness. However, polynomially bounded Rényi divergence *does* imply that any event

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

occurring with negligible probability when X_0 is chosen according to the second distribution also occurs with negligible probability when X_0 is chosen according to the first distribution. For these reasons, we change our security goal from an "indistinguishability-based" one (namely, requiring that, given the transcript, the real session key is indistinguishable from uniform) to an "unpredictability-based" one (namely, given the transcript, it should be infeasible to compute the real session key). In the end, though, once the parties agree on an unpredictable value k they can hash it to obtain the final session key $\mathsf{sk} = H(k)$; this final value sk will be indistinguishable from uniform if H is modeled as a random oracle.

2 Preliminaries

2.1 Notation

Let $\mathbb{Z}$ be the ring of integers, and let $[N] = \{0, 1, \ldots, N-1\}$. If χ is a probability distribution over some set S, then $x_0, x_1, \ldots, x_{\ell-1} \leftarrow \chi$ denotes independently sampling each x_i from distribution χ. We let $\mathrm{Supp}(\chi) = \{x : \chi(x) \neq 0\}$. Given an event E, we use $\overline{E}$ to denote its complement. Let $\chi(E)$ denote the probability that event E occurs under distribution χ. Given a polynomial p_i, let $(p_i)_j$ denote the jth coefficient of p_i. Let $\log(X)$ denote $\log_2(X)$, and $\exp(X)$ denote e^X.

2.2 Ring Learning with Errors

Informally, the (decisional) version of the Ring Learning with Errors (Ring-LWE) problem is: for some secret ring element s, distinguish many random "noisy ring products" with s from elements drawn uniform from the ring. More precisely, the Ring-LWE problem is parameterized by (R, q, χ, ℓ) as follows:

1. R is a ring, typically written as a polynomial quotient ring $R = \mathbb{Z}[X]/(f(X))$ for some irreducible polynomial $f(X)$ in the indeterminate X. In this paper, we restrict to the case of that $f(X) = X^n + 1$ where n is a power of 2. In later sections, we let R be parameterized by n.
2. q is a modulus defining the quotient ring $R_q := R/qR = \mathbb{Z}_q[X]/(f(X))$. We restrict to the case that q is prime and $q = 1 \mod 2n$.
3. $\chi = (\chi_s, \chi_e)$ is a pair of noise distributions over R_q (with χ_s the *secret key* distribution and χ_e the *error* distribution) that are concentrated on "short" elements, for an appropriate definition of "short" (e.g., the Euclidean distance metric on the integer-coefficients of the polynomials s or e drawn from R_q); and
4. ℓ is the number of samples provided to the adversary.

Formally, the Ring-LWE problem is to distinguish between ℓ samples independently drawn from one of two distributions. The first distribution is generated by fixing a random secret $s \leftarrow \chi_s$ then outputting

$$(a_i, b_i = s \cdot a_i + e_i) \in R_q \times R_q,$$

for $i \in [\ell]$, where each $a_i \in R_q$ is drawn uniformly at random and each $e_i \leftarrow \chi_e$ is drawn from the error distribution. For the second distribution, each sample $(a_i, b_i) \in R_q \times R_q$ is simply drawn uniformly at random.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

Let A_{n,q,χ_s,χ_e} be the distribution that outputs the Ring-LWE sample $(a_i, b_i = s \cdot a_i + e_i)$ as above. We denote by $\mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_s,\chi_e,\ell}(\mathcal{B})$ the advantage of algorithm $\mathcal{B}$ in distinguishing distributions $A^{\ell}_{n,q,\chi_s,\chi_e}$ and $\mathcal{U}^{\ell}(R_q^2)$.

We define $\mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_s,\chi_e,\ell}(t)$ to be the maximum advantage of any adversary running in time t. Note that in later sections, we write as $\mathsf{Adv}_{n,q,\chi,\ell}$ when $\chi = \chi_s = \chi_e$ for simplicity.

The Ring-LWE Noise Distribution. The noise distribution χ (here we assume $\chi_s = \chi_e$, though this is not necessary) is usually a discrete Gaussian distribution on $R_q^{\vee}$ or in our case R_q (see [18] for details of the distinction, especially for concrete implementation purposes). Formally, in case of power of two cyclotomic rings, the discrete Gaussian distribution can be sampled by drawing each coefficient independently from the 1-dimensional discrete Gaussian distribution over $\mathbb{Z}$ with parameter σ, which is supported on $\{x \in \mathbb{Z} : -q/2 \leq x \leq q/2\}$ and has density function

$$D_{\mathbb{Z}_q,\sigma}(x) = \frac{e^{\frac{-\pi x^2}{\sigma^2}}}{\sum_{x=-\infty}^{\infty} e^{\frac{-\pi x^2}{\sigma^2}}}.$$

2.3 Rényi divergence

The Rényi divergence (RD) is a measure of closeness of two probability distributions. For any two discrete probability distributions P and Q such that $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$, we define the Rényi divergence of order 2 as

$$\mathrm{RD}_2(P||Q) = \sum_{x \in \mathrm{Supp}(P)} \frac{P(x)^2}{Q(x)}.$$

Rényi divergence has a probability preservation property that can be considered the multiplicative analogues of statistical distance.

Proposition 1. *Given discrete distributions P and Q with $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$, let $E \in \mathrm{Supp}(Q)$ be an arbitrary event. We have*

$$Q(E) \geq P(E)^2 / \mathrm{RD}_2(P||Q).$$

This property implies that as long as $\mathrm{RD}_2(P||Q)$ is bounded by $\mathrm{poly}(\lambda)$, any event E that occurs with negligible probability $Q(E)$ under distribution Q also occurs with negligible probability $P(E)$ under distribution P. We refer to [27, 26] for the formal proof.

Theorem 2.1 ([7]). *Fix $m, q \in \mathbb{Z}$, a bound B, and the 1-dimensional discrete Gaussian distribution $D_{\mathbb{Z}_q,\sigma}$ with parameter σ such that $B < \sigma < q$. Moreover, let $e \in \mathbb{Z}$ be such that $|e| \leq B$. If $\sigma = \Omega(B\sqrt{m/\log \lambda})$, then*

$$\mathrm{RD}_2((e + D_{\mathbb{Z}_q,\sigma})^m || D^m_{\mathbb{Z}_q,\sigma}) \leq \exp(2\pi m (B/\sigma)^2) = \mathrm{poly}(\lambda),$$

where X^m denotes m independent samples from X.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

2.4 Generic Key Reconciliation Mechanism

In this subsection, we define a generic, one round, two-party key reconciliation mechanism which allows both parties to derive the same key from an approximately agreed upon ring element. A key reconciliation mechanism KeyRec consists of two algorithms recMsg and recKey, parameterized by security parameter 1^λ as well as β_{Rec}. In this context, Alice and Bob hold "close" keys – b_A and b_B, respectively – and wish to generate a shared key k so that $k = k_A = k_B$. The abstract mechanism KeyRec is defined as follows:

1. Bob computes $(K, k_B) = \mathsf{recMsg}(b_B)$ and sends the reconciliation message K to Alice.
2. Once receiving K, Alice computes $k_A = \mathsf{recKey}(b_A, K) \in \{0,1\}^\lambda$.

Correctness. Given $b_A, b_B \in R_q$, if each coefficient of $b_B - b_A$ is bounded by β_{Rec} – namely, $|b_B - b_A| \leq \beta_{\mathsf{Rec}}$ – then it is guaranteed that $k_A = k_B$.

Security. A key reconciliation mechanism KeyRec is secure if the subsequent two distribution ensembles are computationally indistinguishable. (First, we describe a simple, helper distribution.)

$\mathsf{Exe}_{\mathsf{KeyRec}}(\lambda)$: A draw from this helper distribution is performed by initiating the key reconciliation protocol among two honest parties and outputting (K, k_B); i.e. the reconciliation message K and (Bob's) key k_B of the protocol execution.

We denote by $\mathsf{Adv}_{\mathsf{KeyRec}}(\mathcal{B})$ the advantage of adversary $\mathcal{B}$ distinguishing the distributions below.

$$\{(K, k_B) : b_B \leftarrow \mathcal{U}(R_q), (K, k_B) \leftarrow \mathsf{Exe}_{\mathsf{KeyRec}}(\lambda, b_B)\}_{\lambda \in \mathbb{N}},$$
$$\{(K, k') : b_B \leftarrow \mathcal{U}(R_q), (K, k_B) \leftarrow \mathsf{Exe}_{\mathsf{KeyRec}}(\lambda, b_B), k' \leftarrow U_\lambda\}_{\lambda \in \mathbb{N}},$$

where U_λ denotes the uniform distribution over λ bits.

We define $\mathsf{Adv}_{\mathsf{KeyRec}}(t)$ to be the maximum advantage of any adversary running in time t.

Key reconciliation mechanisms from the literature. The notion of key reconciliation was first introduced by Ding et al. [19]. in his work on two-party, lattice-based key exchange. It was later used in several important works on two-party key exchange, including [28, 32, 4].

In the key reconciliation mechanisms of Peikert [28], Zhang et al. [32] and Alkim et al. [4], the initiating party sends a small amount of information about its secret, b_B, to the other party. This information is enough to allow the two parties to agree upon the same key $k = k_A = k_B$, while revealing no information about k to an eavesdropper. When instantiating our GKE protocol with this type of key reconciliation (specifically, one of [28, 32, 4]), our final GKE protocol is "contributory," in the sense that all parties contribute entropy towards determining the final key.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

Another method for the two parties to agree upon the same joint key $k = k_A = k_B$, given that they start with keys b_A, b_B that are "close," was first introduced in [3] (we refer to their technique as a key reconciliation mechanism, although it is technically not referred to as such in the literature). Here, the initiating party uses its private input to generate a Regev-style encryption of a random bit string k_B of its choice under secret key b_B. and then sends to the other party, who decrypts with its approximate secret key b_A to obtain k_A. Due to the inherent robustness to noise of Regev-style encryption, it is guaranteed that $k = k_A = k_B$ with all but negligible probability. Instantiating our GKE protocol with this type of key reconciliation (specifically, that in [3]) is also possible, but does not lead to the preferred "contributory GKE," since the initiating party's entropy completely determines the final group key.

3 Group Key Exchange Security Model

A group key-exchange protocol allows a session key to be established among $N > 2$ parties. Following prior work [23, 14, 12, 13], we will use the term group key exchange (GKE) to denote a protocol secure against a *passive* (eavesdropping) adversary and will use the term *group authenticated key exchange* (GAKE) to denote a protocol secure against an *active* adversary, who controls all communication channels. Fortunately, the work of Katz and Yung [23] presents a compiler that takes any GKE protocol and transforms it into a GAKE protocol. The underlying tool required for this transform is any digital signature scheme which is strongly unforgeable under adaptive chosen message attack (EUF-CMA). We may thus focus our attention on achieving GKE in the remainder of this work.

In GKE, the adversary gets to see a single transcript generated by an execution of the GKE protocol. Given the transcript, the adversary must distinguish the real key from a fake key that is generated uniformly at random and independently of the transcript.

Formally, for security parameter $\lambda \in \mathbb{N}$, we define the following distribution:

$\mathsf{Execute}^{\mathcal{O}_H}_{\Pi}(\lambda)$: A draw from this distribution is performed by sampling a classical random oracle $\mathcal{H}$ from distribution $\mathcal{O}_H$, initiating the GKE protocol Π among N honest parties with security parameter λ relative to $\mathcal{H}$, and outputting $(\mathsf{trans}, \mathsf{sk})$—the transcript trans and key sk of the protocol execution.

Consider the following distributions:

$$\{(\mathsf{trans}, \mathsf{sk}) : (\mathsf{trans}, \mathsf{sk}) \leftarrow \mathsf{Execute}^{\mathcal{O}_H}_{\Pi}(\lambda)\}_{\lambda \in \mathbb{N}},$$
$$\{(\mathsf{trans}, \mathsf{sk}') : (\mathsf{trans}, \mathsf{sk}) \leftarrow \mathsf{Execute}^{\mathcal{O}_H}_{\Pi}(\lambda), \mathsf{sk}' \leftarrow U_\lambda\}_{\lambda \in \mathbb{N}},$$

where U_λ denotes the uniform distribution over λ bits. Let $\mathsf{Adv}^{\mathsf{GKE},\mathcal{O}_H}(\mathcal{A})$ denote the advantage of adversary $\mathcal{A}$, with classical access to the sampled oracle $\mathcal{H}$, distinguishing the distributions above.

To enable a concrete security analysis, we define $\mathsf{Adv}^{\mathsf{GKE},\mathcal{O}_H}(t, q_{\mathcal{O}_H})$ to be the maximum advantage of any adversary running in time t and making at most $q_{\mathcal{O}_H}$

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

queries to the random oracle. Security holds even if the adversary sees multiple executions by a hybrid argument.

In the next section we will define our GKE scheme and prove that it satisfies the notion of GKE.

4 A Group Key-Exchange Protocol

In this section, we present our group key exchange construction, GKE, which runs key reconciliation protocol KeyRec as a subroutine. Let KeyRec be parametrized by β_{Rec}. The protocol has two security parameters λ and ρ. λ is the computational security parameter, which is used in the security proof. ρ is the statistical security parameter, which is used in the correctness proof. σ_1, σ_2 are parameters of discrete Gaussian distributions. In this setting, N players $P_0, \ldots, P_{N-1}$ plan to generate a shared session key. The players' indices are taken modulo N.

The structure of the protocol is as follows: All parties agree on "close" keys $b_0 \approx \cdots \approx b_{N-1}$ after the second round. Player $N-1$ then initiates a key reconciliation protocol to allow all users to agree on the same key $k = k_0 = \cdots = k_{N-1}$. Since we are only able to prove that k is difficult to compute for an eavesdropping adversary (but may not be indistinguishable from random), we hash k using random oracle $\mathcal{H}$ to get the final shared key sk.

Public parameter: $R_q = \mathbb{Z}_q[x]/(x^n+1)$, $a \leftarrow \mathcal{U}(R_q)$.

Round 1: Each player P_i samples $s_i, e_i \leftarrow \chi_{\sigma_1}$ and broadcasts $z_i = as_i + e_i$.
Round 2: Player P_0 samples $e'_0 \leftarrow \chi_{\sigma_2}$ and each of the other players P_i samples $e'_i \leftarrow \chi_{\sigma_1}$, broadcasts $X_i = (z_{i+1} - z_{i-1})s_i + e'_i$.
Round 3: Player P_{N-1} proceeds as follows:

1. Samples $e''_{N-1} \leftarrow \chi_{\sigma_1}$ and computes $b_{N-1} = z_{N-2}Ns_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + X_0 \cdot (N-2) + \cdots + X_{N-3}$.
2. Computes $(K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1})$ and broadcasts K_{N-1}.
3. Obtains session key $\mathsf{sk}_{N-1} = \mathcal{H}(k_{N-1})$.

Key Computation: Each player P_i (except P_{N-1}) proceeds as follows:

1. Computes $b_i = z_{i-1}Ns_i + X_i \cdot (N-1) + X_{i+1} \cdot (N-2) + \cdots + X_{i+N-2}$.
2. Computes $k_i = \mathsf{recKey}(b_i, K_{N-1})$, and obtains session key $\mathsf{sk}_i = \mathcal{H}(k_i)$.

4.1 Correctness

The following claim states that each party derives the same session key sk_i, with all but negligible probability, as long as $\chi_{\sigma_1}, \chi_{\sigma_2}$ satisfy the constraint $(N^2+2N) \cdot \sqrt{n}\rho^{3/2}\sigma_1^2 + (\frac{N^2}{2}+1)\sigma_1 + (N-2)\sigma_2 \leq \beta_{\mathsf{Rec}}$, where β_{Rec} is the parameter from the KeyRec protocol.

Theorem 4.1. *Given β_{Rec} as the parameter of KeyRec protocol, $N, n, \rho, \sigma_1, \sigma_2$ as parameters of GKE protocol Π, as long as $(N^2+2N) \cdot \sqrt{n}\rho^{3/2}\sigma_1^2 + (\frac{N^2}{2}+1)\sigma_1 + (N-2)\sigma_2 \leq \beta_{\mathsf{Rec}}$ is satisfied, if all players honestly execute the group key exchange protocol described above, then each player derives the same key as input of $\mathcal{H}$ with probability $1 - 2 \cdot 2^{-\rho}$.*

Proof. We refer to Section A of Appendix for the detailed proof. □

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

5 Security Proof

The following theorem shows that the protocol Π is a passively secure group key-exchange protocol in random oracle model based on Ring-LWE assumption.

Theorem 5.1. *If the parameters in the group key exchange protocol Π satisfy the constraints $2N\sqrt{n}\lambda^{3/2}\sigma_1^2 + (N-1)\sigma_1 \leq \beta_{\mathsf{R\acute{e}nyi}}$ and $\sigma_2 = \Omega(\beta_{\mathsf{R\acute{e}nyi}}\sqrt{n/\log\lambda})$, and if $\mathcal{H}$ is modeled as a classical random oracle, then for any algorithm $\mathcal{A}$ running in time t, making at most q queries to the random oracle, the maximum advantage of $\mathcal{A}$ in breaking GKE security is as follows:*

$$\mathsf{Adv}_{\Pi}^{\mathsf{GKE},\mathcal{O}_H}(t,\mathsf{q}) \leq 2^{-\lambda+1} + \sqrt{\left(N \cdot \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1) + \mathsf{Adv}_{\mathsf{KeyRec}}(t_2) + \frac{\mathsf{q}}{2^{\lambda}}\right) \cdot \frac{\exp\left(2\pi n\left(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2\right)^2\right)}{1-2^{-\lambda+1}}},$$

where $t_1 = t + \mathcal{O}(N)\cdot t_{\mathsf{ring}}, t_2 = t + \mathcal{O}(N)\cdot t_{\mathsf{ring}}$ and where t_{ring} is defined as the (maximum) time required to perform operations in R_q.

Proof. Consider the joint distribution of (T,sk), where $\mathsf{T} = (\{z_i\},\{X_i\},K_{k-1})$ is the transcript of an execution of the protocol Π, and k is the final shared session key. The distribution of (T,sk) is denoted as Real. Proceeding via a sequence of experiments, we will show that under the Ring-LWE assumption, if an efficient adversary queries the random oracle on input k_{N-1} in the Ideal experiment (to be formally defined) with at most negligible probability, then it also queries the random oracle on input k_{N-1} in the Real experiment with at most negligible probability.

Furthermore, in Ideal, the input k_{N-1} to the random oracle is uniform random, which means that the adversary has $\mathsf{negl}(\lambda)$ probability of guessing k_{N-1} in Ideal when $\mathsf{q} = \mathrm{poly}(\lambda)$. Finally, we argue that the above is sufficient to prove the GKE security of the scheme, because in the random oracle model, the output of the random oracle on k_{N-1} – i.e. the agreed upon key – looks uniformly random to an adversary who does not query k_{N-1}. We now proceed with the formal proof.

Let Query be the event that k_{N-1} is among the adversary $\mathcal{A}$'s random oracle queries and denote by $\Pr_i[\mathsf{Query}]$ the probability that event Query happens in *Experiment i*. Note that we let $e_0' = \hat{e}_0$ in order to distinguish this from the other e_i''s sampled from a different distribution.

Experiment 0. This is the original experiment. In this experiment, the distribution of (T,sk) is as follows, denoted Real :

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

$$\mathsf{Real} := \left\{ \left(\begin{array}{l} a \leftarrow R_q; s_0, s_1, \ldots, s_{N-1}, e_0, e_1, \ldots, e_{N-1} \leftarrow \chi; \\ z_0 = as_0 + e_0, z_1 = as_1 + e_1, \ldots, z_{N-1} = as_{N-1} + e_{N-1}; \\ e'_1, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}; \hat{e}_0 \leftarrow \chi_{\sigma_2}; \\ X_0 = (z_1 - z_{N-1})s_0 + \hat{e}_0, X_1 = (z_2 - z_0)s_1 + e'_1, \ldots, \\ X_{N-1} = (z_0 - z_{N-2})s_{N-1} + e'_{N-1}; e''_{N-1} \leftarrow \chi_{\sigma_1}; \\ b_{N-1} = z_{N-2}Ns_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right) : (\mathsf{T}, \mathsf{sk}) \right\}$$

Since $\Pr[\mathcal{A}\text{succeeds}] = \frac{1}{2} + \mathsf{Adv}_{\Pi}^{\mathsf{GKE},\mathcal{O}_H}(t, \mathsf{q}) = \Pr_0[\mathsf{Query}] \cdot 1 + \Pr_0(\overline{\mathsf{Query}}) \cdot \frac{1}{2}$, we have

$$\mathsf{Adv}_{\Pi}^{\mathsf{GKE},\mathcal{O}_H}(t, \mathsf{q}) \leq \Pr_0[\mathsf{Query}]. \tag{1}$$

In the remainder of the proof, we focus on bounding $\Pr_0[\mathsf{Query}]$.

Experiment 1. In this experiment, X_0 is replaced by $X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0$. The remainder of the experiment is exactly the same as *Experiment 0*. The corresponding distribution of $(\mathsf{T}, \mathsf{sk})$ is as follows, denoted Dist_1:

$$\mathsf{Dist}_1 := \left\{ \left(\begin{array}{l} a \leftarrow \mathcal{U}(R_q); s_0, s_1, \ldots, s_{N-1}, e_0, e_1, \ldots, e_{N-1} \leftarrow \chi_{\sigma_1}; \\ z_0 = as_0 + e_0, z_1 = as_1 + e_1, \ldots, z_{N-1} = as_{N-1} + e_{N-1}; \\ e_0, e'_1, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}; \hat{e}_0 \leftarrow \chi_{\sigma_2} \\ X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1 = (z_2 - z_0)s_1 + e'_1, \ldots, \\ X_{N-1} = (z_0 - z_{N-2})s_{N-1} + e'_{N-1}; e''_{N-1} \leftarrow \chi_{\sigma_1}; \\ b_{N-1} = z_{N-2}Ns_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right) : (\mathsf{T}, \mathsf{sk}) \right\}$$

Claim. Given $a \leftarrow \mathcal{U}(R_q)$, $s_0, s_1, \ldots, s_{N-1}, e_0, e_1, \ldots, e_{N-1}, e'_1, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}$, $\hat{e}_0 \leftarrow \chi_{\sigma_2}$, $X_0 = (z_1 - z_{N-1})s_0 + \hat{e}_0$, $X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0$, where $R_q, \chi_{\sigma_1}, \chi_{\sigma_2}$, $z_1, z_{N-1}, X_1, \ldots, X_{N-1}$ are defined as above, and the constraint $2N\sqrt{n}\lambda^{3/2}\sigma_1^2 + (N-1)\sigma_1 \leq \beta_{\mathsf{R\acute{e}nyi}}$ is satisfied, we have

$$\Pr_0[\mathsf{Query}] \leq \sqrt{\left(\Pr_1[\mathsf{Query}] \cdot \frac{\exp(2\pi n(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2)^2)}{1 - 2^{-\lambda+1}} + 2^{-\lambda+1} \right.}. \tag{2}$$

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

Proof. Let $\mathsf{Error} = \sum_{i=0}^{N-1}(s_i e_{i+1} + s_i e_{i-1}) + \sum_{i=1}^{N-1} e'_i$. We begin by showing that the absolute value of each coefficient of Error is bounded by $\beta_{\mathsf{R\acute{e}nyi}}$ with all but negligible probability. Then by adding a "bigger" error $\hat{e}_0 \leftarrow \chi_{\sigma_2}$, the small difference between distributions $\mathsf{Error} + \chi_{\sigma_2}$ (corresponding to Experiment 0) and χ_{σ_2} (corresponding to Experiment 1) can be "washed" away by applying Theorem 2.1.

For all coefficient indices j, note that $|\mathsf{Error}_j| = |(\sum_{i=0}^{N-1}(s_i e_{i+1} + s_i e_{i-1}) + \sum_{i=1}^{N-1} e'_i)_j|$. Let bound_λ denote the event that for all i and all coordinate indices j, $|(s_i)_j| \leq c\sigma_1$, $|(e_i)_j| \leq c\sigma_1$, $|(e'_i)_j| \leq c\sigma_1$,$|(e''_{N-1})_j| \leq c\sigma_1$, and $|(\hat{e}_0)_j| \leq c\sigma_2$, where $c = \sqrt{\frac{2\lambda}{\pi \log e}}$. By replacing ρ with λ in Lemma A.1 and Lemma A.2 and by a union bound, we have – conditioned on bound_λ – that $|\mathsf{Error}_j| \leq 2N\sqrt{n}\lambda^{3/2}\sigma_1^2 + (N-1)\sigma_1$ for all j, with probability at least $1 - 2N \cdot 2n2^{-2\lambda}$. Since, under the assumption that $4Nn \leq 2^\lambda$, we have that $\Pr[\mathsf{bound}_\lambda] \geq 1 - 2^{-\lambda}$, we conclude that

$$\Pr[|\mathsf{Error}_j| \leq \beta_{\mathsf{R\acute{e}nyi}}, \forall j] \geq 1 - 2^{-\lambda+1}. \tag{3}$$

For a fixed $\mathsf{Error} \in R_q$, we denote by D_1 the distribution of $\mathsf{Error} + \chi_{\sigma_2}$ and note that D_1, χ_{σ_2} are n-dimension distributions.

Since $\sigma_2 = \Omega(\beta_{\mathsf{R\acute{e}nyi}}\sqrt{n/\log\lambda})$, assuming that for all j, $|\mathsf{Error}_j| \leq \beta_{\mathsf{R\acute{e}nyi}}$, by Theorem 2.1, we have

$$\mathrm{RD}_2(D_1||\chi_{\sigma_2}) \leq \exp(2\pi n(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2)^2) = \mathrm{poly}(\lambda). \tag{4}$$

Then it is straightforward to verify that the distribution of X_0 in *Experiment 0* is

$$as_1 s_0 - as_{N-1}s_0 - \sum_{i=0}^{N-1}(e_{i+1}s_i + e_{i-1}s_i) - \sum_{i=1}^{N-1} e'_i + D_1,$$

and the distribution of X'_0 in *Experiment 1* is

$$as_1 s_0 - as_{N-1}s_0 - \sum_{i=0}^{N-1}(e_{i+1}s_i + e_{i-1}s_i) - \sum_{i=1}^{N-1} e'_i + \chi_{\sigma_2}.$$

In addition, the remaining part of Dist_1 is identical to Real. Therefore we may view Real in *Experiment 0* as a function of a random variable sampled from D_1 and take Dist_1 in *Experiment 1* as a function of a random variable sampled from χ_{σ_2}.

Recall that Query is the event that k_{N-1} is contained in the set of random oracle queries issued by adversary $\mathcal{A}$. We denote by Xbound the event that $|\mathsf{Error}_j| \leq \beta_{\mathsf{R\acute{e}nyi}}, \forall j$. Note that computation of Error_j is available in both *Experiment 0* and *Experiment 1*. We denote by $\Pr_0[\mathsf{Xbound}]$ (resp. $\Pr_1[\mathsf{Xbound}]$) the probability that event Xbound occurs in *Experiment 0* (resp. *Experiment 1*) and define $\Pr_0[\overline{\mathsf{Xbound}}], \Pr_1[\overline{\mathsf{Xbound}}]$ analogously. Let Real' (resp. Dist'_1) denote the random variable Real (resp. Dist_1), conditioned on the event Xbound. Therefore,

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

we have

$$\begin{aligned}
\Pr_0[\mathsf{Query}] &= \Pr_0[\mathsf{Query}|\mathsf{Xbound}] \cdot \Pr_0[\mathsf{Xbound}] + \Pr_0[\mathsf{Query}|\overline{\mathsf{Xbound}}] \cdot \Pr_0[\overline{\mathsf{Xbound}}] \\
&\leq \Pr_0[\mathsf{Query}|\mathsf{Xbound}] + \Pr_0[\overline{\mathsf{Xbound}}] \\
&\leq \Pr_0[\mathsf{Query}|\mathsf{Xbound}] + 2^{-\lambda+1} \\
&\leq \sqrt{\Pr_1[\mathsf{Query}|\mathsf{Xbound}] \cdot \mathrm{RD}_2(\mathsf{Real}'||\mathsf{Dist}_1')} + 2^{-\lambda+1} \\
&\leq \sqrt{\Pr_1[\mathsf{Query}|\mathsf{Xbound}] \cdot \mathrm{RD}_2(D_1||\chi_{\sigma_2})} + 2^{-\lambda+1} \\
&\leq \sqrt{\Pr_1[\mathsf{Query}|\mathsf{Xbound}] \cdot \exp(2\pi n(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2)^2)} + 2^{-\lambda+1} \\
&\leq \sqrt{\Pr_1[\mathsf{Query} \cdot \frac{\exp(2\pi n(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2)^2)}{\Pr_1[\mathsf{Xbound}]}} + 2^{-\lambda+1} \\
&\leq \sqrt{\Pr_1[\mathsf{Query}] \cdot \frac{\exp(2\pi n(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2)^2)}{1-2^{-\lambda+1}}} + 2^{-\lambda+1},
\end{aligned}$$

where the second and last inequalities follow from (3), the third inequality follows from Proposition 1 and the fifth inequality follows from (4). □

In Section B of the Appendix, we show that

$$\Pr_1[\mathsf{Query}] \leq \left(N \cdot \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1) + \mathsf{Adv}_{\mathsf{KeyRec}}(t_2) + \frac{\mathsf{q}}{2^\lambda}\right)\Bigg($$

which concludes the proof of Theorem 5.1. □

5.1 Parameter Constraints

Beyond the parameter settings recommended for instantiating Ring-LWE with security parameter λ, parameters $N, n, \sigma_1, \sigma_2, \lambda, \rho$ of the protocol above are also required to satisfy the following inequalities:

$$(N^2+2N) \cdot \sqrt{n}\rho^{3/2}\sigma_1^2 + (\frac{N^2}{2}+1)\sigma_1 + (N-2)\sigma_2 \leq \beta_{\mathsf{Rec}} \quad \text{(Correctness)} \quad (5)$$

$$2N\sqrt{n}\lambda^{3/2}\sigma_1^2 + (N-1)\sigma_1 \leq \beta_{\mathsf{R\acute{e}nyi}} \quad \text{(Security)} \quad (6)$$

$$\sigma_2 = \Omega(\beta_{\mathsf{R\acute{e}nyi}}\sqrt{n/\log\lambda}) \quad \text{(Security)} \quad (7)$$

We comment that once the ring, the noise distributions, and the security parameters λ, ρ are fixed, the maximum number of parties is fixed.

6 Acknowledgments

This material is based on work performed under financial assistance award 70NANB15H328 from the U.S. Department of Commerce, National Institute of Standards and Technology. Work by Dana Dachman-Soled was additionally supported in part by NSF grants #CNS-1840893 and #CNS-1453045, and by a research partnership award from Cisco.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

References

1. Michel Abdalla, Emmanuel Bresson, Olivier Chevassut, and David Pointcheval. Password-based group key exchange in a constant number of rounds. In *9th Intl. Conference on Theory and Practice of Public Key Cryptography (PKC)*, volume 3958 of *Lecture Notes in Computer Science*, pages 427–442. Springer, 2006.
2. Michel Abdalla and David Pointcheval. A scalable password-based group key exchange protocol in the standard model. In *Advances in Cryptology—Asiacrypt 2006*, volume 4284 of *Lecture Notes in Computer Science*, pages 332–347. Springer, 2006.
3. Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. NewHope without reconciliation. Cryptology ePrint Archive, Report 2016/1157, 2016. http://eprint.iacr.org/2016/1157.
4. Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. Post-quantum key exchange—a new hope. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 327–343, Austin, TX, 2016. USENIX Association.
5. Klaus Becker and Uta Wille. Communication complexity of group key distribution. In *Proceedings of the 5th ACM Conference on Computer and Communications Security*, CCS '98, pages 1–6, New York, NY, USA, 1998.
6. Mihir Bellare and Phillip Rogaway. Provably secure session key distribution: The three party case. In *27th Annual ACM Symposium on Theory of Computing*, pages 57–66, Las Vegas, NV, USA, May 29 – June 1, 1995. ACM Press.
7. Andrej Bogdanov, Siyao Guo, Daniel Masny, Silas Richelson, and Alon Rosen. On the hardness of learning with rounding over small modulus. In *Theory of Cryptography Conference*, pages 209–224. Springer, 2016.
8. Jens-Matthias Bohli, Maria Isabel Gonzalez Vasco, and Rainer Steinwandt. Password-authenticated constant-round group key establishment with a common reference string. Cryptology ePrint Archive, Report 2006/214, 2006. http://eprint.iacr.org/2006/214.
9. Jens-Matthias Bohli, María Isabel González Vasco, and Rainer Steinwandt. Secure group key establishment revisited. *International Journal of Information Security*, 6(4):243–254, Jul 2007.
10. Dan Boneh, Darren Glass, Daniel Krashen, Kristin Lauter, Shahed Sharif, Alice Silverberg, Mehdi Tibouchi, and Mark Zhandry. Multiparty non-interactive key exchange and more from isogenies on elliptic curves. *arXiv preprint arXiv:1807.03038*, 2018.
11. Emmanuel Bresson and Dario Catalano. Constant round authenticated group key agreement via distributed computation. In Feng Bao, Robert Deng, and Jianying Zhou, editors, *PKC 2004: 7th Intl. Workshop on Theory and Practice in Public Key Cryptography*, volume 2947 of *Lecture Notes in Computer Science*, pages 115–129, Singapore, March 1–4, 2004. Springer.
12. Emmanuel Bresson, Olivier Chevassut, and David Pointcheval. Provably authenticated group Diffie-Hellman key exchange – the dynamic case. In Colin Boyd, editor, *Advances in Cryptology—Asiacrypt 2001*, volume 2248 of *Lecture Notes in Computer Science*, pages 290–309, Gold Coast, Australia, December 9–13, 2001. Springer.
13. Emmanuel Bresson, Olivier Chevassut, and David Pointcheval. Dynamic group Diffie-Hellman key exchange under standard assumptions. In Lars R. Knudsen, editor, *Advances in Cryptology—Eurocrypt 2002*, volume 2332 of *Lecture Notes in Computer Science*, pages 321–336, Amsterdam, The Netherlands, April 28 – May 2, 2002. Springer.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

14. Emmanuel Bresson, Olivier Chevassut, David Pointcheval, and Jean-Jacques Quisquater. Provably authenticated group Diffie-Hellman key exchange. In *ACM CCS 01: 8th Conference on Computer and Communications Security*, pages 255–264, Philadelphia, PA, USA, November 5–8, 2001. ACM Press.
15. Mike Burmester and Yvo Desmedt. A secure and efficient conference key distribution system (extended abstract). In Alfredo De Santis, editor, *Advances in Cryptology—Eurocrypt'94*, volume 950 of *Lecture Notes in Computer Science*, pages 275–286. Springer, 1995.
16. Mike Burmester and Yvo Desmedt. A secure and scalable group key exchange system. *Information Processing Letters*, 94(3):137–143, May 2005.
17. Kyu Young Choi, Jung Yeon Hwang, and Dong Hoon Lee. Efficient ID-based group key agreement with bilinear maps. In Feng Bao, Robert Deng, and Jianying Zhou, editors, *PKC 2004: 7th Intl. Workshop on Theory and Practice in Public Key Cryptography*, volume 2947 of *Lecture Notes in Computer Science*, pages 130–144, Singapore, March 1–4, 2004. Springer.
18. Eric Crockett and Chris Peikert. Challenges for ring-LWE. Cryptology ePrint Archive, Report 2016/782, 2016. http://eprint.iacr.org/2016/782.
19. Jintai Ding, Xiang Xie, and Xiaodong Lin. A simple provably secure key exchange scheme based on the learning with errors problem. Cryptology ePrint Archive, Report 2012/688, 2012. http://eprint.iacr.org/2012/688.
20. Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.
21. I. Ingemarsson, D. Tang, and C. Wong. A conference key distribution system. *IEEE Trans. Inf. Theor.*, 28(5):714–720, September 1982.
22. Jonathan Katz and Ji Sun Shin. Modeling insider attacks on group key-exchange protocols. In *Proceedings of the 12th ACM Conference on Computer and Communications Security*, CCS '05, pages 180–189, New York, NY, USA, 2005. ACM.
23. Jonathan Katz and Moti Yung. Scalable protocols for authenticated group key exchange. In Dan Boneh, editor, *Advances in Cryptology—Crypto 2003*, volume 2729 of *Lecture Notes in Computer Science*, pages 110–125, Santa Barbara, CA, USA, 2003. Springer.
24. Jonathan Katz and Moti Yung. Scalable protocols for authenticated group key exchange. *Journal of Cryptology*, 20(1):85–113, 2007.
25. Yongdae Kim, Adrian Perrig, and Gene Tsudik. Simple and fault-tolerant key agreement for dynamic collaborative groups. In *Proceedings of the 7th ACM Conference on Computer and Communications Security*, CCS '00, pages 235–244, New York, NY, USA, 2000.
26. Adeline Langlois, Damien Stehlé, and Ron Steinfeld. GGHLite: More efficient multilinear maps from ideal lattices. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology—Eurocrypt 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 239–256, Copenhagen, Denmark, May 11–15, 2014. Springer.
27. Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In Henri Gilbert, editor, *Advances in Cryptology—Eurocrypt 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 1–23, French Riviera, May 30 – June 3, 2010. Springer.
28. Chris Peikert. Lattice cryptography for the internet. Cryptology ePrint Archive, Report 2014/070, 2014. http://eprint.iacr.org/2014/070.
29. D. G. Steer and L. Strawczynski. A secure audio teleconference system. In *MILCOM 88, 21st Century Military Communications - What's Possible?'. Conference record. Military Communications Conference*, Oct 1988.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

30. M. Steiner, G. Tsudik, and M. Waidner. Key agreement in dynamic peer groups. *IEEE Transactions on Parallel and Distributed Systems*, 11(8):769–780, Aug 2000.
31. Qianhong Wu, Yi Mu, Willy Susilo, Bo Qin, and Josep Domingo-Ferrer. Asymmetric group key agreement. In Antoine Joux, editor, *Advances in Cryptology—Eurocrypt 2009*, volume 5479 of *Lecture Notes in Computer Science*, pages 153–170, Cologne, Germany, April 26–30, 2009. Springer.
32. Jiang Zhang, Zhenfeng Zhang, Jintai Ding, Michael Snook, and Özgür Dagdelen. Authenticated key exchange from ideal lattices. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology—Eurocrypt 2015, Part II*, volume 9057 of *Lecture Notes in Computer Science*, pages 719–751, Sofia, Bulgaria, April 26–30, 2015. Springer.

A Correctness of the Group Key-Exchange Protocol

Theorem 4.1. *Given β_{Rec} as parameter of* KeyRec *protocol, $N, n, \rho, \sigma_1, \sigma_2$ as parameters of GKE protocol Π, $(N^2+2N)\cdot\sqrt{n}\rho^{3/2}\sigma_1^2+(\frac{N^2}{2}+1)\sigma_1+(N-2)\sigma_2 \leq \beta_{\mathsf{Rec}}$ is satisfied, if all players honestly execute the group key exchange protocol as described above, then each player derive the same key as input of $\mathcal{H}$ with probability $1-2\cdot 2^{-\rho}$.*

Proof. Given $s_i, e_i, e'_i, e''_{N-1} \leftarrow \chi_{\sigma_1}$, $\hat{e}_0 \leftarrow \chi_{\sigma_2}$ for all i as specified in protocol Π, we begin by introducing the following lemmas to analyze probabilities that each coordinate of $s_i, e_i, e'_i, e''_{N-1}, \hat{e}_0$ are "short" for all i, and conditioned on the first event, $s_i e_i$ are "short".

Lemma A.1. *Given $s_i, e_i, e'_i, e''_{N-1}, \hat{e}_0$ for all i as defined above, let* bound *denote the event that for all i and all coordinate indices j, $|(s_i)_j| \leq c\sigma_1$, $|(e_i)_j| \leq c\sigma_1$, $|(e'_i)_j| \leq c\sigma_1$,$|(e''_{N-1})_j| \leq c\sigma_1$, and $|(\hat{e}_0)_j| \leq c\sigma_2$, where $c = \sqrt{\frac{2\rho}{\pi \log e}}$, we have $\Pr[\mathsf{bound}] \geq 1-2^{-\rho}$.*

Proof. Using the fact that complementary error function $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt \leq e^{-x^2}$, we obtain

$$\begin{aligned}\Pr[|v| \geq c\sigma+1; v \leftarrow D_{\mathbb{Z}_q,\sigma}] &\leq 2\sum_{x=\lfloor c\sigma+1\rfloor}^{\infty} D_{\mathbb{Z}_q,\sigma}(x) \leq \frac{2}{\sigma}\int_{c\sigma}^{\infty} e^{-\frac{\pi x^2}{\sigma^2}}dx \\ &= \frac{2}{\sqrt{\pi}}\int_{\frac{\sqrt{\pi}}{\sigma}(c\sigma)}^{\infty} e^{-t^2}dt \leq e^{-c^2\pi}.\end{aligned}$$

Note that there are $3nN$ number of coordinates sampled from distribution $D_{\mathbb{Z}_q,\sigma_1}$, and n number of coordinates sampled from distribution $D_{\mathbb{Z}_q,\sigma_2}$ in total. Assume $3nN+n \leq e^{c^2\pi/2}$, since all the coordinates are sampled independently, we bound $\Pr[\mathsf{bound}]$ as follow:

$$\begin{aligned}\Pr[\mathsf{bound}] &= \Big(1 - \Pr[|v| \geq c\sigma_1+1; v \leftarrow D_{\mathbb{Z}_q,\sigma_1}]\Big)^{3nN} \\ &\quad \cdot \big(1-\Pr[|\hat{e}_0| \geq c\sigma_2+1; \hat{e}_0 \leftarrow D_{\mathbb{Z}_q,\sigma_2}]\big)^n \\ &\geq 1-(3nN+n)e^{-c^2\pi} \geq 1-e^{-c^2\pi/2} \geq 1-2^{-\rho}.\end{aligned}$$

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

The last inequality follows as $c = \sqrt{\frac{2\rho}{\pi \log e}}$. □

Lemma A.2. *Given $s_i, e_i, e'_i, e''_{N-1}, \hat{e}_0$ for all i as defined above, and* bound *as defined in Lemma A.1, let* $\mathsf{product}_{\mathsf{s_i},\mathsf{e_j}}$ *denote the event that, for all coefficient indices v, $|(s_i e_j)_v| \leq \sqrt{n}\rho^{3/2}\sigma_1^2$. we have*

$$\Pr[\mathsf{product}_{\mathsf{s_i},\mathsf{e_j}} | \mathsf{bound}] \geq 1 - 2n \cdot 2^{-2\rho}.$$

Proof. For $t \in \{0, \ldots, n-1\}$, Let $(s_i)_t$ denote the t^{th} coefficient of $s_i \in R_q$, namely, $s_i = \sum_{t=0}^{n-1}(s_i)_t X^i$. $(e_j)_t$ is defined analogously. Since we have $X^n + 1$ as modulo of R, it is easy to see that $(s_i e_j)_v = c_v X^v$, where $c_v = \sum_{u=0}^{n-1}(s_i)_u (e_j)^*_{v-u}$, and $(e_j)^*_{v-u} = (e_j)_{v-u}$ if $v - u \geq 0$, $(e_j)^*_{v-u} = -(e_j)_{v-u+n}$, otherwise. Thus, conditioned on $|(s_i)_t| \leq c\sigma_1$ and $|(e_j)_t| \leq c\sigma_1$ (for all i, j, t) where $c = \sqrt{\frac{2\rho}{\pi \log e}}$, by Hoeffding's Inequality [20], we derive

$$\Pr[|(s_i e_j)_v| \geq \delta] = \Pr\left[\left|\sum_{u=0}^{n-1}(s_i)_u (e_j)^*_{v-u}\right| \geq \delta\right] \leq 2\exp\left(\frac{-2\delta^2}{n(2c^2\sigma_1^2)^2}\right),$$

as each product $(s_i)_u (e_j)^*_{v-u}$ in the sum is an independent random variable with mean 0 in the range $[-c^2\sigma_1^2, c^2\sigma_1^2]$. By setting $\delta = \sqrt{n}\rho^{3/2}\sigma_1^2$, we obtain

$$\Pr[|(s_u e_v)_i| \geq \sqrt{n}\rho^{3/2}\sigma_1^2] \leq 2^{-2\rho+1}. \tag{8}$$

Finally, by Union Bound,

$$\Pr[\mathsf{product}_{\mathsf{s_i},\mathsf{e_j}} | \mathsf{bound}] = \Pr[|(s_i e_j)_v| \leq \sqrt{n}\rho^{3/2}\sigma_1^2, \forall v] \geq 1 - 2n \cdot 2^{-2\rho}. \tag{9}$$

□

Now we begin analyzing the chance that not all parties agree on the same final key. The correctness of KeyRec guarantees that this group key exchange protocol has agreed session key among all parties $\forall i, k_i = k_{N-1}$, if $\forall j$, the j^{th} coefficient of $|b_{N-1} - b_i| \leq \beta_{\mathsf{Rec}}$.

For better illustration, we first write $X_0, \ldots, X_{N-1}$ in form of linear system as follows. $\boldsymbol{X} = [X_0 \;\; X_1 \;\; X_2 \;\; \cdots \;\; X_{N-1}]^T$

$$= \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & -1 \\ -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} as_0 s_1 \\ as_1 s_2 \\ as_2 s_3 \\ as_3 s_4 \\ \vdots \\ as_{N-2} s_{N-1} \\ as_{N-1} s_0 \end{bmatrix}}_{\boldsymbol{S}} + \underbrace{\begin{bmatrix} s_0 e_1 - s_0 e_{N-1} + e'_0 \\ s_1 e_2 - s_1 e_0 + e'_1 \\ s_2 e_3 - s_2 e_1 + e'_2 \\ s_3 e_4 - s_3 e_2 + e'_3 \\ \vdots \\ s_{N-2} e_{N-3} - s_{N-2} e_{N-3} + e'_{N-2} \\ s_{N-1} e_0 - s_{N-1} e_{N-2} + e'_{N-1} \end{bmatrix}}_{\boldsymbol{E}}. \tag{10}$$

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

We denote the matrices above by $\boldsymbol{M}, \boldsymbol{S}, \boldsymbol{E}$ from left to right and have the linear system as $\boldsymbol{X} = \boldsymbol{MS}+\boldsymbol{E}$. By setting $\boldsymbol{B}_i = [i-1 \;\; i-2 \;\; \cdots \;\; 0 \;\; N-1 \;\; N-2 \;\; \cdots \;\; i]$ as a N-dimensional vector, we can then write b_i as $\boldsymbol{B}_i \cdot \boldsymbol{X} + N(as_is_{i-1}+s_ie_{i-1}) = \boldsymbol{B}_i\boldsymbol{MS}+\boldsymbol{B}_i\boldsymbol{E}+N(as_is_{i-1}+s_ie_{i-1})$, for $i \neq N-1$ and write b_{N-1} as $\boldsymbol{B}_{N-1}\boldsymbol{MS}+\boldsymbol{B}_{N-1}\boldsymbol{E}+N(as_{N-1}s_{N-2}+s_{N-1}e_{N-2})+e''_{N-1}$. It is straightforward to see that, entries of $\boldsymbol{MS}$ and Nas_is_{i-1} are eliminated through the process of computing $b_{N-1}-b_i$. Thus we get

$$
\begin{aligned}
b_{N-1} - b_i &= (\boldsymbol{B}_{N-1} - \boldsymbol{B}_i)\,\boldsymbol{E} + N(s_{N-1}e_{N-2} - s_ie_{i-1}) + e''_{N-1} \\
&= (N-i-1)\cdot\left(\left(\sum_{j\in\mathbb{Z}\cap[0,i-1] \text{ and } j=N-1}\left(s_je_{j+1} - s_je_{j-1} + e'_j\right)\right) + e''_{N-1}\right) \\
&\quad + (-i-1)\left(\sum_{j=i}^{N-2}\left(s_je_{j+1} - s_je_{j-1} + e'_j\right)\right) + N(s_{N-1}e_{N-2} - s_ie_{i-1})
\end{aligned}
$$

Observe that for an arbitrary $i \in [N]$, there are at most (N^2+2N) terms in form of s_ue_v, at most $N^2/2$ terms in form of e'_w where $e'_w \leftarrow \chi_{\sigma_1}$, at most $N-2$ terms of e'_0, where $e'_0 \leftarrow \chi_{\sigma_2}$, and one term in form of e''_{N-1} in any coordinate of the sum above. Let $\mathsf{product}_{\mathsf{ALL}}$ denote the event that for all the terms in form of s_ue_v observed above, each coefficient of such term is bounded by $\sqrt{n}\rho^{3/2}\sigma_1^2$. By Union Bound and by assuming $2n(N^2+2N) \leq 2^\rho$, it is straightforward to see $\Pr[\overline{\mathsf{product}_{\mathsf{ALL}}}|\mathsf{bound}] \leq (N^2+2N)\cdot 2n2^{-2\rho} \leq 2^{-\rho}$.

Let bad be the event that not all parties agree on the same final key. Given the constraint $(N^2+2N)\cdot\sqrt{n}\rho^{3/2}\sigma_1^2 + (\frac{N^2}{2}+1)\sigma_1 + (N-2)\sigma_2 \leq \beta_{\mathsf{Rec}}$ satisfied, we have

$$
\begin{aligned}
\Pr[\mathsf{bad}] &= \Pr[\mathsf{bad}|\mathsf{bound}]\cdot\Pr[\mathsf{bound}] + \Pr[\mathsf{bad}|\overline{\mathsf{bound}}]\cdot\Pr[\overline{\mathsf{bound}}] && (11)\\
&\leq \Pr[\overline{\mathsf{product}_{\mathsf{ALL}}}]\cdot 1 + 1\cdot\Pr[\overline{\mathsf{bound}}] \leq 2\cdot 2^{-\rho}, && (12)
\end{aligned}
$$

which completes the proof.

□

B Concluding the Proof of Theorem 5.1

Theorem 5.1 (Restated)**.** *If the parameters in group key exchange protocol Π satisfy the constraints that $2N\sqrt{n}\lambda^{3/2}\sigma_1^2+(N-1)\sigma_1 \leq \beta_{\mathsf{Rényi}}, \sigma_2 = \Omega(\beta_{\mathsf{Rényi}}\sqrt{n/\log\lambda})$, and $\mathcal{H}$ is modeled as a classical random oracle, then for any algorithm $\mathcal{A}$ running in time t, making at most q queries to the random oracle, the maximum advantage of $\mathcal{A}$ in breaking GKE security is as follows:*

$$
\begin{aligned}
&\mathsf{Adv}_{\Pi}^{\mathsf{GKE},\mathcal{O}_H}(t,\mathsf{q}) \leq 2^{-\lambda+1} \\
&+\sqrt{\left(N\cdot\mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1) + \mathsf{Adv}_{\mathsf{KeyRec}}(t_2) + \frac{\mathsf{q}}{2^\lambda}\right)\cdot\frac{\exp\left(2\pi n\left(\beta_{\mathsf{Rényi}}/\sigma_2\right)^2\right)}{1-2^{-\lambda+1}}},
\end{aligned}
$$

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

where t_1 and t_2 equal to $t+\mathcal{O}(N)\cdot t_{\mathsf{ring}}$ and t_{ring} is the time to perform operations in R_q.

Proof. (Continued) Recall that *Experiment 0* is the real world experiment. We have that $\mathsf{Adv}_{\Pi}^{\mathsf{GKE},\mathcal{O}_H}(t,\mathsf{q}) \leq \Pr_0[\mathsf{Query}]$ (see Equation 1), where Query is the event that k_{N-1} is among the adversary $\mathcal{A}$'s random oracle queries and $\Pr_i[\mathsf{Query}]$ is the probability that event Query happens in *Experiment i*.

In *Experiment 1*, we switched from X_0 as sampled in the real world to $X_0' = -\sum_{i=1}^{N-1} X_i + \hat{e}_0$ and showed (see Equation 2) that

$$\Pr_0[\mathsf{Query}] \leq \sqrt{\Pr_1[\mathsf{Query}] \cdot \frac{\exp(2\pi n(\beta_{\mathsf{R\acute{e}nyi}}/\sigma_2)^2)}{1-2^{-\lambda+1}}} + 2^{-\lambda+1}.$$

Therefore, to prove the theorem, it remains to show that

$$\Pr_1[\mathsf{Query}] \leq \left(N \cdot \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1) + \mathsf{Adv}_{\mathsf{KeyRec}}(t_2) + \frac{\mathsf{q}}{2^{\lambda}}\right).$$

We do so by considering a sequence of experiments as follows:

Experiment 2. This experiment proceeds exactly the same as *Experiment 1*, except that z_0 is generated uniformly at random, instead of being generated as an Ring-LWE instance. The corresponding distribution is as follows, denoted Dist_2:

$$\mathsf{Dist}_2 := \left\{ \begin{array}{l} a \leftarrow \mathcal{U}(R_q); s_1,\ldots,s_{N-1},e_1,\ldots,e_{N-1} \leftarrow \chi_{\sigma_1}; \\ z_0 \leftarrow \mathcal{U}(R_q), z_1 = as_1+e_1,\ldots,z_{N-1}=as_{N-1}+e_{N-1}; \\ e_1',\ldots,e_{N-1}' \leftarrow \chi_{\sigma_1}; \hat{e}_0 \leftarrow \chi_{\sigma_2} \\ X_0' = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1 = (z_2-z_0)s_1+e_2',\ldots, \\ X_{N-1} = (z_0-z_{N-2})s_{N-1}+e_{N-1}'; e_{N-1}'' \leftarrow \chi_{\sigma_1}; \\ b_{N-1} = z_{N-2}Ns_{N-1}+e_{N-1}''+X_{N-1}\cdot(N-1)+ \\ \qquad X_0\cdot(N-2)+\cdots+X_{N-3}; \\ (K_{N-1},k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0,\ldots,z_{N-1},X_0,\ldots,X_{N-1},K_{N-1}). \end{array} : (\mathsf{T},\mathsf{sk}) \right\}$$

Bounding the difference of $|\Pr_2[\mathsf{Query}] - \Pr_1[\mathsf{Query}]|$:

Given algorithm $\mathcal{A}$ running in time t attacking Π, let $\mathcal{B}$ be an algorithm running in time t_1 that takes as input (a,z_0), generates (T,sk) based on distribution Dist_1' which is identical to Dist_1 except for (a,z_0) given as input, runs $\mathcal{A}$ as subroutine and outputs whatever $\mathcal{A}$ outputs. It is straightforward to see that if (a,z_0) is sampled from the Ring-LWE distribution $A_{n,q,\chi_{\sigma_1}}$, then Dist_1' is identical to Dist_1, and if (a,z_0) is sampled from $\mathcal{U}(R_q^2)$, Dist_1' is identical to Dist_2. Note that t_1 is equal to t plus a minor overhead for the simulation of the security experiment for $\mathcal{A}$.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

Therefore we conclude that the difference of algorithm $\mathcal{A}$'s success probability in *Experiment 1* and *Experiment 2* is bounded by probability that $\mathcal{B}$ running in time t_1 distinguishes $A_{n,q,\chi_{\sigma_1}}$ from $\mathcal{U}(R_q)$ given one sample. Since $\mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1) \geq \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},2}(t_1) \geq \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},1}(t_1)$, for simplicity, we have

$$|\Pr_2[\mathsf{Query}] - \Pr_1[\mathsf{Query}]| \leq \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1). \tag{13}$$

Recall that in the previous experiment, we switched z_0 to be uniformly distributed in R_q. In next two experiments, we switch z_1, X_1 to be elements uniformly distributed in R_q.

Experiment 3. the experiment proceeds exactly the same as *Experiment 2*, except for setting $z_0 = z_2 - r_1, X_1 = r_1 s_1 + e'_1$, where r_1 is sampled from $\mathcal{U}(R_q)$. The corresponding distribution is as follows, denoted as Dist_3.

Bounding the difference of $|\Pr_3[\mathsf{Query}] - \Pr_2[\mathsf{Query}]|$: Since r_1 is sampled uniformly, $z_2 - r_1$ is also a uniformly distributed random value, then we claim that *Experiment 3* is identical to *Experiment 3* up to variable substitution, namely

$$\Pr_3[\mathsf{Query}] = \Pr_2[\mathsf{Query}]. \tag{14}$$

$$\mathsf{Dist}_3 := \left\{ \begin{array}{l} a \leftarrow \mathcal{U}(R_q), r_1 \leftarrow \mathcal{U}(R_q); \\ s_1, \ldots, s_{N-1}, e_1, \ldots, e_{N-1} \leftarrow \chi_{\sigma_1}; z_0 = z_2 - r_1, \\ z_1 = as_1 + e_1, \ldots, z_{N-1} = as_{N-1} + e_{N-1}; \\ e'_1, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}; \hat{e}_0 \leftarrow \chi_{\sigma_2}; X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, \\ \left(\begin{array}{l} X_1 = r_1 s_1 + e'_1, X_2 = (z_3 - z_1)s_2 + e'_2, \\ \ldots, X_{N-1} = (z_0 - z_{N-2})s_{N-1} + e'_{N-1}; \end{array} \right. : (\mathsf{T}, \mathsf{sk}) \\ e''_{N-1} \leftarrow \chi_{\sigma_1}; \\ b_{N-1} = z_{N-2} N s_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right) \left(\right.$$

Experiment 4. This experiment proceeds exactly the same as *Experiment 3*, except that z_1, X_1 are uniformly distributed in R_q. The corresponding distribution is as follows, denoted as Dist_4.

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

$$\mathsf{Dist}_4 := \left\{ \begin{array}{l} \Big(a, r_1 \leftarrow \mathcal{U}(R_q); s_2, \ldots, s_{N-1}, e_2, \ldots, e_{N-1} \leftarrow \chi_{\sigma_1}; \\ z_0 = z_2 - r_1, z_1 \leftarrow \mathcal{U}(R_q), z_2 = as_2 + e_2, \ldots, \\ z_{N-1} = as_{N-1} + e_{N-1}; e'_2, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}; \\ \hat{e}_0 \leftarrow \chi_{\sigma_2}; X'_0 = -\sum_{i=1}^{N-1} \Big(X_i + \hat{e}_0, \\ \Big(X_1 \leftarrow \mathcal{U}(R_q), X_2 = (z_3 - z_1)s_2 + e'_2, \qquad : (\mathsf{T}, \mathsf{sk}) \Big\} \Big(\\ \ldots, X_{N-1} = (z_0 - z_{N-2})s_{N-1} + e'_{N-1}; e''_{N-1} \leftarrow \chi_{\sigma_1}; \\ b_{N-1} = z_{N-2} N s_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right) \Big($$

Bounding the difference of $|\Pr_4[\mathsf{Query}] - \Pr_3[\mathsf{Query}]|$:

Given an algorithm $\mathcal{A}$ running in time t attacking Π, let $\mathcal{B}$ be an algorithm running in time t_1 that takes as input (a, z_1), (r_1, X_1), generates $(\mathsf{T}, \mathsf{sk})$ based on distribution Dist'_3 which is identical to Dist_3 except for (a, z_1), (r_1, X_1) given as input. $\mathcal{B}$ runs $\mathcal{A}$ as a subroutine and outputs whatever $\mathcal{A}$ outputs. Note that t_1 is equal to t plus a minor overhead for the simulation of the security experiment for $\mathcal{A}$.

It is clear to see that if (a, z_1) and (r_1, X_1) are sampled from the Ring-LWE distribution $A_{n,q,\chi_{\sigma_1}}$, then Dist'_3 is identical to Dist_3. If (a, z_1) and (r_1, X_1) are sampled from $\mathcal{U}(R_q^2)$, Dist'_3 is identical to Dist_4.

Therefore we conclude that the difference of algorithm $\mathcal{A}$ successful probability in winning *Experiment 4* and *Experiment 3* is bounded by the advantage of adversary $\mathcal{B}$ running in time t_1 in distinguishing $A_{n,q,\chi_{\sigma_1}}$ from $\mathcal{U}(R_q^2)$ given two samples. Thus,

$$|\Pr_4[\mathsf{Query}] - \Pr_3[\mathsf{Query}]| \leq \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1). \tag{15}$$

Experiment 5. This experiment proceeds exactly the same as *Experiment 4*, except that z_0 is sampled directly from $\mathcal{U}(R_q)$. We leave the formal definition of Dist_5 implicit for simplicity.

Bounding the difference of $|\Pr_5[\mathsf{Query}] - \Pr_4[\mathsf{Query}]|$: It is easy to see that the corresponding distribution Dist_5 is identical to Dist_4 by substituting variable z_0 for $z_2 - r_1$. Thus,

$$\Pr_5[\mathsf{Query}] = \Pr_4[\mathsf{Query}]. \tag{16}$$

In the case that $N \geq 3$, we present the following sequence of experiments from *Experiment 6* to *Experiment* $3N - 4$. For $i = 2, 3, \ldots, N-2$, we define three experiments *Experiment* $3i$, *Experiment* $3i + 1$, *Experiment* $3i + 2$. It is ensured that in the experiments prior to *Experiment* $3i$, we already switched

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

z_j, X_j for all $0 \leq j \leq i-1$. In *Experiment* $3i$, *Experiment* $3i+1$ and *Experiment* $3i+2$, we replace z_i and X_i by random elements uniformly distributed in R_q. *Experiment* $3i$, *Experiment* $3i+1$, *Experiment* $3i+2$ are formally defined as follows:

Experiment $3i$**.** The experiment proceeds exactly the same as *Experiment* $3i-1$, except for setting $z_{i-1} = z_{i+1} - r_i, X_i = r_i s_i + e'_i$, where r_1 is sampled from $\mathcal{U}(R_q)$. The corresponding distribution is as follows, denoted Dist_{3i}

$$\mathsf{Dist}_{3i} := \left\{ \begin{array}{l} a, r_i \leftarrow \mathcal{U}(R_q); s_i, \ldots, s_{N-1}, e_i, \ldots, e_{N-1} \leftarrow \chi_{\sigma_1}; \\ z_0, \ldots, z_{i-2} \leftarrow \mathcal{U}(R_q), z_{i-1} = z_{i+1} - r_i, z_i = a s_i + e_i, \\ \ldots, z_{N-1} = a s_{N-1} + e_{N-1}; e'_i, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}; \\ \hat{e}_0 \leftarrow \chi_{\sigma_2}; X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1, \ldots, X_{i-1} \leftarrow \mathcal{U}(R_q) \quad : (\mathsf{T}, \mathsf{sk}) \\ \left(\begin{array}{l} X_i = r_i s_i + e'_i, X_{i+1} = (z_{i+2} - z_i) s_{i+1} + e'_{i+1}, \\ \ldots, X_{N-1} = (z_0 - z_{N-2}) s_{N-1} + e'_{N-1}; e''_{N-1} \leftarrow \chi_{\sigma_1}; \end{array} \right) \\ b_{N-1} = z_{N-2} N s_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right)$$

Experiment $3i+1$**.** This experiment proceeds exactly the same as *Experiment* $3i$, except that z_i, X_i are uniformly distributed in R_q. The corresponding distribution is as follows, denoted Dist_{3i+1}:

$$\mathsf{Dist}_{3i+1} := \left\{ \begin{array}{l} a, r_i \leftarrow \mathcal{U}(R_q); s_{i+1}, \ldots, s_{N-1}, e_{i+1}, \ldots, e_{N-1} \leftarrow \chi_{\sigma_1} \\ z_0, \ldots, z_{i-2} \leftarrow \mathcal{U}(R_q), z_{i-1} = z_{i+1} - r_i, z_i \leftarrow \mathcal{U}(R_q), \\ z_{i+1} = a s_{i+1} + e_{i+1}, \ldots, z_{N-1} = a s_{N-1} + e_{N-1}; \\ e'_1, \ldots, e'_{N-1} \leftarrow \chi_{\sigma_1}; \hat{e}_0 \leftarrow \chi_{\sigma_2} \\ X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1, \ldots, X_{i-1} \leftarrow \mathcal{U}(R_q), \quad : (\mathsf{T}, \mathsf{sk}) \\ \left(\begin{array}{l} X_i \leftarrow \mathcal{U}(R_q), X_{i+1} = (z_{i+2} - z_i) s_{i+1} + e'_{i+1}, \ldots, \\ X_{N-1} = (z_0 - z_{N-2}) s_{N-1} + e'_{N-1}; e''_{N-1} \leftarrow \chi_{\sigma_1}; \end{array} \right) \\ b_{N-1} = z_{N-2} N s_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right)$$

Experiment $3i+2$**.** This experiment proceeds exactly the same as *Experiment* $3i+1$, except that z_{i-1} is directly sampled from $\mathcal{U}(R_q)$. The corresponding distribution is denoted as Dist_{3i+2}. We leave the formal definition of Dist_{3i+2} implicit

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

for simplicity.

Bounding the difference of $|\Pr_{3i}[\mathsf{Query}]-\Pr_{3i-1}[\mathsf{Query}]|$, $|\Pr_{3i+1}[\mathsf{Query}]-\Pr_{3i}[\mathsf{Query}]|$, and $|\Pr_{3i+2}[\mathsf{Query}]-\Pr_{3i+1}[\mathsf{Query}]|$ follows exactly the same logic as *bounding the differences of* $|\Pr_3[\mathsf{Query}]-\Pr_2[\mathsf{Query}]|$, $|\Pr_4[\mathsf{Query}]-\Pr_3[\mathsf{Query}]|$, and $|\Pr_5[\mathsf{Query}]-\Pr_4[\mathsf{Query}]|$, respectively. Then we have

$$\Pr_{3i}[\mathsf{Query}] = \Pr_{3i-1}[\mathsf{Query}]; \tag{17}$$

$$|\Pr_{3i+1}[\mathsf{Query}] - \Pr_{3i}[\mathsf{Query}]| \leq \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1); \tag{18}$$

$$\Pr_{3i+2}[\mathsf{Query}] = \Pr_{3i+1}[\mathsf{Query}]; \tag{19}$$

Note that in *Experiment* $3N-4$, the last experiment of the experiment sequence above, we already switched all the z_i, X_i up to z_{N-1}, X_{N-1}. We construct the next two experiments to switch $z_{N-1}, X_{N-1}, b_{N-1}$.

Experiment $3N-3$**.** The experiment proceeds exactly the same as *Experiment* $3N-4$, except that we let $z_{N-2} = r_2, X_{N-1} = r_1 s_{N-1} + e'_{N-1}, z_0 = r_1 + r_2$, where r_1, r_2 are uniformly distributed in R_q. The corresponding distribution is as follows, denoted Dist_{3N-3}.

Bounding the difference of $|\Pr_{3N-3}[\mathsf{Query}] - \Pr_{3N-4}[\mathsf{Query}]|$:

Since r_1, r_2 is sampled uniformly, $r_1 + r_2$ is also uniformly distributed in R_q. Then we claim that *Experiment* $3N-3$ is identical to *Experiment* $3N-4$ up to variable substitution, written as

$$\Pr_{3N-3}[\mathsf{Query}] = \Pr_{3N-4}[\mathsf{Query}]; \tag{20}$$

$$\mathsf{Dist}_{3N-3} := \left\{ \begin{array}{l} \Big(q, r_1, r_2 \leftarrow \mathcal{U}(R_q), s_{N-1}, e_{N-1} \leftarrow \chi_{\sigma_1}; z_0 = r_1 + r_2, \\ z_1, \ldots, z_{N-3} \leftarrow \mathcal{U}(R_q), z_{N-2} = r_2, \\ X_{N-1} = a s_{N-1} + e_{N-1}; \hat{e}_0 \leftarrow \chi_{\sigma_2}; e'_{N-1} \leftarrow \chi_{\sigma_1}; \\ X'_0 = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1, \ldots, X_{N-2} \leftarrow \mathcal{U}(R_q), \\ X_{N-1} = r_1 s_{N-1} + e'_{N-1}; e''_{N-1} \leftarrow \chi_{\sigma_1}; \\ b_{N-1} = r_2 N s_{N-1} + e''_{N-1} + X_{N-1} \cdot (N-1) + \\ \qquad X_0 \cdot (N-2) + \cdots + X_{N-3}; \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right\} \Big(: (\mathsf{T}, \mathsf{sk}) \Big)$$

Experiment $3N-2$**.** This experiment proceeds exactly the same as *Experiment* $3N-3$, except that $z_{N-1}, X_{N-1}, b_{N-1}$ are generated from $\mathcal{U}(R_q)$. The

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

corresponding distribution is as follows, denoted Dist_{3N-2}:

$$\mathsf{Dist}_{3N-2} := \left\{ \begin{array}{l} a \leftarrow \mathcal{U}(R_q), z_0, z_1 \ldots, z_{N-2} \leftarrow \mathcal{U}(R_q), \\ z_{N-1} \leftarrow \mathcal{U}(R_q); \hat{e}_0 \leftarrow \chi_{\sigma_2}; r_1, r_2 \leftarrow \mathcal{U}(R_q) \\ X_0' = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1, \ldots, X_{N-1} \leftarrow \mathcal{U}(R_q) \qquad : (\mathsf{T}, \mathsf{sk}) \\ b_{N-1} \leftarrow \mathcal{U}(R_q); \\ (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}); \mathsf{sk} = \mathcal{H}(k_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X_0, \ldots, X_{N-1}, K_{N-1}). \end{array} \right\}$$

Bounding the difference of $|\Pr_{3N-2}[\mathsf{Query}] - \Pr_{3N-3}[\mathsf{Query}]|$:

Let $b_{rlwe} = r_2 N s_{N-1} + e''_{N-1}$, then $b_{N-1} = b_{rlwe} + X_{N-1} \cdot (N-1) + X_0 \cdot (N-2) + \cdots + X_{N-3}$. As r_2 is sampled uniformly at random and N is invertible over R_q, $r_2 N$ is uniformly distributed in R_q.

Given an algorithm $\mathcal{A}$ running in time t attacking group key exchange protocol Π, let $\mathcal{B}$ be an algorithm that takes as input (a, z_{N-1}), (r_1, X_{N-1}), and $(r_2 N, b_{rlwe})$, generates $(\mathsf{T}, \mathsf{sk})$ based on distribution Dist'_{3N-3} which is identical to Dist_{3N-3} except for (a, z_{N-1}), (r_1, X_{N-1}), and $(r_2 N, b_{rlwe})$ given as input. $\mathcal{B}$ runs $\mathcal{A}$ as subroutine and outputs whatever $\mathcal{A}$ outputs. Note that running time t_1 of $\mathcal{B}$ equals to t plus a minor overhead for the simulation of the security experiment for $\mathcal{A}$.

It is straightforward to see that if (a, z_{N-1}), (r_1, X_1), and $(r_2 N, b_{rlwe})$ are sampled from the Ring-LWE distribution $A_{n,q,\chi_{\sigma_1}}$, then Dist'_{3N-3} is identical to Dist_{3N-3}. If (a, z_{N-1}), (r_1, X_{N-1}), and $(r_2 N, b_{rlwe})$ are sampled from $\mathcal{U}(R_q^2)$, then Dist'_{3N-3} is identical to Dist_{3N-2}, since when b_{rlwe} is sampled uniformly at random, $b_{rlwe} + X_{N-1} \cdot (N-1) + X_0 \cdot (N-2) + \cdots + X_{N-3}$ is also uniformly distributed over R_q.

Therefore we conclude that the difference of algorithm $\mathcal{A}^{\mathsf{GKE}}$'s success probability in *Experiment 3N - 2* and *Experiment 3N - 3* is bounded by the advantage of adversary $\mathcal{B}$ running in time t_1 in distinguishing Ring-LWE from $\mathcal{U}(R_q)$ given three samples. Thus, we conclude that

$$|\Pr_{3N-2}[\mathsf{Query}] - \Pr_{3N-3}[\mathsf{Query}]| \leq \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1). \tag{21}$$

Experiment $3N-1$**.** This experiment proceeds exactly the same as *Experiment* $3N-2$, except that k_{N-1} is directly sampled uniformly from $\{0,1\}^\lambda$. Note that the corresponding distribution is exactly the distribution Ideal.

$$\mathsf{Ideal} := \left\{ \begin{array}{l} a \leftarrow \mathcal{U}(R_q); z_0, \ldots, z_{N-1} \leftarrow \mathcal{U}(R_q); e'_0 \leftarrow \chi_{\sigma_1}; \\ X_0' = -\sum_{i=1}^{N-1} X_i + \hat{e}_0, X_1, \ldots, X_{N-1} \leftarrow \mathcal{U}(R_q) \\ b_{N-1} \leftarrow \mathcal{U}(R_q); (K_{N-1}, k_{N-1}) = \mathsf{recMsg}(b_{N-1}) \quad : (\mathsf{T}, \mathsf{sk}) \\ k'_{N-1} \leftarrow \{0,1\}^\lambda; \mathsf{sk} = \mathcal{H}(k'_{N-1}); \\ \mathsf{T} = (z_0, \ldots, z_{N-1}, X'_0, \ldots, X_{N-1}, K_{N-1}); \end{array} \right\}$$

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

Bounding the difference of $|\Pr_{3N-1}[\mathsf{Query}] - \Pr_{3N-2}[\mathsf{Query}]|$:

Given transcript T, and b_{N-1} which is uniformly distributed, using a straight forward reduction, we obtain advantage of adversary $\mathcal{B}$ running in time t_2 in distinguishing k_{N-1} computed by $\mathsf{recMsg}(b_{N-1})$ from a uniform bit string k'_{N-1} with length λ is at least $|\Pr_{3N-1}[\mathsf{Query}] - \Pr_{3N-2}[\mathsf{Query}]|$, namely,

$$|\Pr_{3N-1}[\mathsf{Query}] - \Pr_{3N-2}[\mathsf{Query}]| \leq \mathsf{Adv}_{\mathsf{KeyRec}}(t_2). \tag{22}$$

Note that t_2 equals to the running time of adversary $\mathcal{A}$ attacking the protocol Π, plus a minor overhead for simulating experiment for $\mathcal{A}$.

Finally, since adversary attacking the GKE protocol Π makes at most q queries to the random oracle, $\Pr_{3N-1}[\mathsf{Query}] = \frac{\mathsf{q}}{2^\lambda} \in \mathsf{negl}(\lambda)$. Combining Equations (13) - (22), we have

$$\Pr_1[\mathsf{Query}] \leq N \cdot \mathsf{Adv}^{\mathsf{RLWE}}_{n,q,\chi_{\sigma_1},3}(t_1) + \mathsf{Adv}_{\mathsf{KeyRec}}(t_2) + \frac{\mathsf{q}}{2^\lambda}. \tag{23}$$

The theorem now follows immediately from Equations (1), (2), and (23). □

Apon, Daniel; Dachman-Soled, Dana; Gong, Huijing; Katz, Jonathan.
"Constant-Round Group Key Exchange from the Ring-LWE Assumption."
Paper presented at The Tenth International Conference on Post-Quantum Cryptography (PQCrypto 2019), Chongqing, China. May 8, 2019 - May 10, 2019.

Magnetotransport in highly enriched ^{28}Si for quantum information processing devices

A. N. Ramanayaka[1,2], Ke Tang[1,2], J. A. Hagmann[1], Hyun-Soo Kim[1,2],
C. A. Richter[1], and J. M. Pomeroy[1]

[1] *National Institute of Standards & Technology, Gaithersburg, Maryland 20899, USA*
[2] *Joint Quantum Institute, University of Maryland, College Park, Maryland 20742, USA*
e-mail: aruna.ramanayaka@nist.gov

Elimination of unpaired nuclear spins can result in low error rates for quantum computation; therefore, isotopically enriched ^{28}Si is regarded as an ideal environment for quantum information processing devices. Using mass selected ion beam deposition technique, we *in-situ* enrich and deposit epitaxial ^{28}Si achieving better than $99.99998\ \%$ ^{28}Si isotope fractions [1]. To explore the electrical properties and optimize the growth conditions of *in-situ* enriched ^{28}Si, we fabricate top gated Hall bar devices, and investigate the magnetotransport in this material at magnetic fields as high as $12\ \mathrm{T}$ and temperatures ranging from $10\ \mathrm{K}$ to $1.2\ \mathrm{K}$. A schematic of the cross-sectional view and an optical micrograph of the fabricated device is shown in fig.1 (a) and (b), respectively. At a temperature $1.9\ \mathrm{K}$, we measure maximum mobilities of approximately $(1740 \pm 2)\ \mathrm{cm}^2/(\mathrm{V}\cdot\mathrm{s})$ and $(6040 \pm 3)\ \mathrm{cm}^2/(\mathrm{V}\cdot\mathrm{s})$ at an electron density of $\approx 1.2 \times 10^{12}\ \mathrm{cm}^{-2}$ for devices fabricated on ^{28}Si and natural Si, respectively. For magnetic fields $B > 2\ \mathrm{T}$, both types of devices demonstrate well developed Shubnikov-de Haas oscillations on the longitudinal magnetoresistance (see fig2 and fig3). In contrast to the device on isotopically enriched ^{28}Si epi-layer, the device on natural Si demonstrates spin-splitting for $B > 3\ \mathrm{T}$ (see fig2). Furthermore, relative to the device on $^{nat:}$Si, the weak-localization is stronger for the device fabricated on isotopically enriched ^{28}Si. Based on the T dependence of the Shubnikov-de Haas oscillations and weak-localization, the dominant scattering mechanism in these devices appears to be background impurity scattering and/or interface roughness scattering. We believe that the relatively lower mobility and stronger weak-localization observed for the devices fabricated on ^{28}Si may be due to the dilute adventitious C, N and O in deposited ^{28}Si.

[1] K. J. Dwyer, J. M. Pomeroy, D. S. Simons, K. L. Steffens, and J. W. Lau, Journal of Physics D: Applied Physics **47**, 345105 (2014).

Ramanayaka, Aruna; Tang, Ke; Hagmann, Joseph; Kim, Hyun; Richter, Curt; Pomeroy, Joshua.
"Magnetotransport in highly enriched 28Si for quantum information processing devices."
Paper presented at 2018 Workshop on Innovative Nanoscale Devices and Systems (WINDS), Kohala Coast, HI, United States. November 25, 2018 - November 30, 2018.

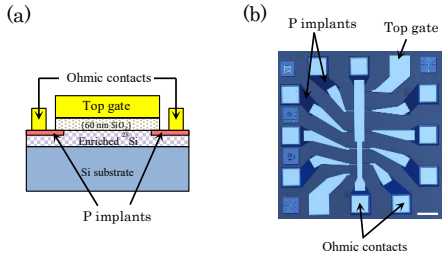


Fig.1: (a) Schematic representations of the gated Hall bar devices fabricated on ^{28}Si and (b) An optical micrograph of a gated multi terminal Hall bar device fabricated on ^{28}Si are shown.

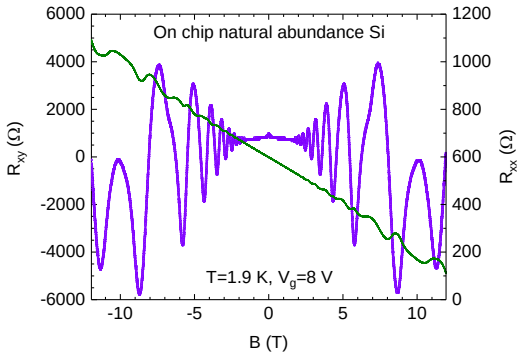


Fig.2: Magnetoresistance R_{xx} (right-axis) and Hall resistance R_{xy} (left-axis) measured for the device fabricated on natural Si substrate are shown. Note that the device on $^{nat.}$Si and ^{28}Si are fabricated on the same Si substrate. Therefore, both have the same processing history.

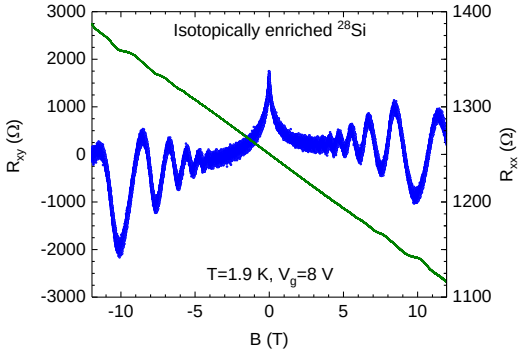


Fig.3: Magnetoresistance R_{xx} (right-axis) and Hall resistance R_{xy} (left-axis) measured for the device fabricated on isotopically enriched ^{28}Si epi-layer are shown.

Ramanayaka, Aruna; Tang, Ke; Hagmann, Joseph; Kim, Hyun; Richter, Curt; Pomeroy, Joshua.
"Magnetotransport in highly enriched 28Si for quantum information processing devices."
Paper presented at 2018 Workshop on Innovative Nanoscale Devices and Systems (WINDS), Kohala Coast, HI, United States. November 25, 2018 - November 30, 2018.

2019 IEEE 37th VLSI Test Symposium (VTS)

RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level

Miao (Tony) He[1], Jungmin Park[1], Adib Nahiyan[1], Apostol Vassilev[2], Yier Jin[1], and Mark Tehranipoor[1]
Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611[1]
National Institute of Standards and Technology, Gaithersburg, MD 20899[2]
Email: {tonyhe, jungminpark, adib1991}ufl.edu, apostol.vassilev@nist.gov, {yier.jin, tehranipoor}@ece.ufl.edu

***Abstract*—Power side-channel attacks (SCAs) have become a major concern to the security community due to their non-invasive feature, low-cost, and effectiveness in extracting secret information from hardware implementation of cryto algorithms. Therefore, it is imperative to evaluate if the hardware is vulnerable to SCAs during its design and validation stages. Currently, however, there is little known effort in evaluating the vulnerability of a hardware to SCAs at early design stage. In this paper, we propose, for the first time, an automated framework, named RTL-PSC, for power side-channel leakage assessment of hardware crypto designs at register-transfer level (RTL) with built-in evaluation metrics. RTL-PSC first estimates power profile of a hardware design using functional simulation at RTL. Then it utilizes the evaluation metrics, comprising of KL divergence metric and the success rate (SR) metric based on maximum likelihood estimation to perform power side-channel leakage (PSC) vulnerability assessment at RTL. We analyze Galois-Field (GF) and Look-up Table (LUT) based AES designs using RTL-PSC and validate its effectiveness and accuracy through both gate-level simulation and FPGA results. RTL-PSC is also capable of identifying blocks* inside the design that contribute the most to the PSC vulnerability which can be used for efficient countermeasure implementation.†**

***Index terms*—** Side-channel Attacks, Leakage Assessment, Vulnerability Evaluation, Register-Transfer Level.

I. Introduction

Power side-channel attacks (SCAs) exploit the weaknesses in the hardware implementations of crypto algorithms to leak sensitive information, e.g., the encryption key, irrespective of the mathematical robustness of the algorithms. A number of SCAs namely Simple Power Analysis [1], Differential Power Analysis (DPA) [1], Correlation Power Analysis (CPA) [2], Template Attacks [3], Mutual Information Analysis (MIA) [4], and Partitioning Power Analysis (PPA) [5] have been proposed and successfully demonstrated over the past two decades. These attacks exploit the fact that the power consumption of a cryptographic hardware depends on the data it processes and the operation it performs [6].

To counter these attacks, a variety of countermeasures have been proposed to make the crypto hardware resilient to SCAs. The goal of these countermeasures is to reduce the dependencies between the intermediate values of the cryptographic algorithms and the power consumption of the cryptographic devices. For example, hiding countermeasures attempt to break the link between the processed data values and the power consumption of the devices [6]. However, all of the SCA countermeasures adversly affect the circuit area, thus making them impractical to be applied to modern resource-constrained designs. One major reason may come from the fact that the evaluation methodology for side-channel leakage assessment are not capable of identifying the source of vulnerabilities effectively and accurately. Hence, the corresponding countermeasure has to be applied to the entire design instead of specific sub-blocks responsible for power side-channel leakage (PSC) vulnerability.

Apart from power side-channel attacks and their corresponding countermeasures, another highly important topic in this domain is power side-channel leakage assessment. Several techniques have been proposed in this domain including signal-to-noise ratio (SNR) [7], Test Vector Leakage Assessment (TVLA) methodology [8], success rate [9], and autonomous side-channel vulnerability evaluator (AMASIVE) [10], [11]. However, these techniques suffer from the following major limitations.

- They mostly focus on the post-silicon side-channel assessment, which is too late and prohibitively expensive in making any changes to the design to address the leakage issue.
- Existing techniques typically require large amount of plaintexts and power traces, hence need prohibitively large simulation time, i.e., these techniques are feasible for the post-silicon stage rather than the pre-silicon stage.
- Some techniques, e.g., TVLA [8] and χ^2-test [12] can only provide a pass/fail test and cannot give quantitative measure of PSC vulnerability which can lead to false positive results.
- Existing techniques mostly require a security analysts to be manually involved in leakage assessment, which may not be feasible due to cost and time-to-market constraints for modern devices.

We summarize the evaluation time and accuracy of side-channel leakage assessment as well as the flexibility to make design changes at different pre-silicon design stages w.r.t. the post-fabricated device level in Figure 1. In the pre-silicon stage, as the leakage assessment accuracy increases, the leakage evaluation time increases exponentially from RTL to gate-level (GTL) to layout level. It can also be observed that the flexibility in making design changes is reduced from RTL to gate-level to layout level. On the other hand, when observing the post-silicon stage, leakage assessment can be performed efficiently and accurately, however, flexibility for making design changes is very difficult (as in FPGA) if not impossible (as in ASIC).

Our Contributions: We propose a framework named RTL-PSC which can automatically assess PSC vulnerability at the earliest pre-silicon design stage, i.e., RTL. This framework is developed to be integrated into the traditional ASIC and FPGA design flow. RTL-PSC provides distinctive capabilities to chip designers and security analysts as listed below:

- **Leakage assessment at early design stage.** The RTL-PSC framework is developed to automatically evaluate PSC vulnerability of a design at higher levels of abstraction to reduce time-to-market, cost of redesign, and the overall cost of adding security to the design.
- **Technology independent.** The vulnerability analysis is performed on RTL design to make the evaluation framework fairly library independent.

*'design' refers to top module as well as its constituting sub-modules, while a 'block' refers to a sub-module within the design.

†DISCLAIMER: This paper is not subject to copyright in the United States. Commercial products are identified in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the identified products are necessarily the best available for the purpose.

He, Miao (Tony); Park, Jungmin; Nahiyan, Adib; Vassilev, Apostol; Jin, Yier; Tehranipoor, Mark.
"RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level."
Paper presented at 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, United States. April 23, 2019 - April 25, 2019.

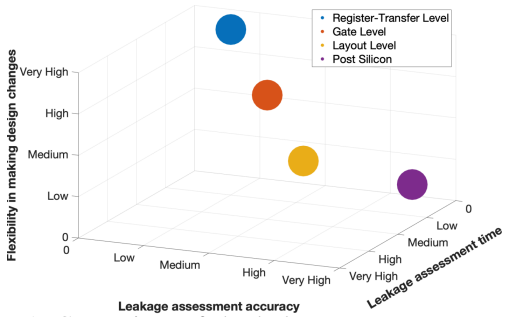


Figure 1: Comparison of the leakage assessment at various stages of the design process.

- **Fine granularity evaluation.** The RTL-PSC framework identifies which block of a design contributes the most to the vulnerability, i.e., pinpoints which block is leaking more secret information. Countermeasures only need to be applied for the vulnerable blocks/modules to reduce area overhead significantly.
- **Comprehensive analysis.** RTL-PSC considers the relation between blocks/modules of a design during vulnerability analysis instead of analyzing blocks/modules of a design independently.
- **Fast power estimation.** RTL-PSC estimates power leakage distribution based on the number of transitions at RTL to ensure the evaluation framework is fast.
- **Generic framework.** The RTL-PSC framework can be automatically applied to any cryptographic implementations at RTL without any customization or designers' involvement.
- **Time, accuracy, and flexibility trade-off.** RTL-PSC can accurately and efficiently estimate PSC vulnerability at RTL. RTL-PSC has an average evaluation time of 35*mins* and its evaluation results closely matches with silicon results obtained from FPGA. Also, RTL-PSC provides the hardware designers with the most flexibility to address PSC vulnerabilities having the capability of working at RTL.

The rest of the paper is organized as follows: In Section II, the RTL-PSC framework is presented. Section III presents the RTL side-channel vulnerability evaluation metrics. Section IV provides and analyzes in detail the results demonstrating the performance of RTL-PSC and its respective metrics proposed in this paper. Finally, concluding remarks are offered in Section V.

II. RTL-PSC Vulnerability Evaluation Framework

Figure 2 outlines the side-channel vulnerability evaluation framework. It includes two main parts, RTL Switching Activity Interchange Format (SAIF) file generation shown in the blue box and identification of vulnerable designs and blocks shown in the purple box. Algorithm 1 describes the identification technique for vulnerable designs and blocks. Note that, here TC refers to the transition count, $\mathcal{N}$ refers to the Gaussian distribution, f refers to the probability density function (PDF), D_{KL} refers to the KL divergence and ML refers to the maximum likelihood. Specifically, in Step 1, a group of simulation keys are specified. In Step 2, we utilize Synopsys VCS to perform functional simulation of the RTL design with the plaintexts and the selected keys as the inputs (key selection process is described in Section III-A). In Step 3, once the simulation is complete, the SAIF file for the RTL design is generated. After finishing Step 3, all SAIF files for a group of keys and the applied plaintexts are generated (See Algorithm 1, Lines 4-8). As its name indicates, the SAIF file includes the switching activity information for each net and register in the RTL design. Moreover, the SAIF file generated based on the RTL design has the same hierarchy as the design itself, hence, in Step 4, the SAIF file for each module in the design can be separated for localized vulnerability analysis. Next, the evaluation metrics are applied for leakage assessment. Specifically, in Step 5, the obtained switching activity is exploited to estimate the power leakage distribution for the design and each module within it (Lines 10-11 in Algorithm 1). In Step 6, the Kullback-Leibler (KL) divergence [13] and success rate (SR) based on power leakage distribution are calculated for the design and each block (Lines 12-15 in Algorithm 1). In Step 7, vulnerability analysis is performed for the design and each block. In Step 8, the vulnerable design is identified based on the analysis performed in the previous step. Then the vulnerable blocks in the design that are leaking information the most are identified for further processing. Step 7 and Step 8 correspond to Lines 16-18 in Algorithm 1. Following this, the framework enters into Step 9, where countermeasures only need to be applied to the vulnerable block(s). Note that Step 9 is outside the scope in this paper.

Algorithm 1 Identifying Vulnerable Blocks

1: **procedure** IDENTIFYING VULNERABLE BLOCKS
2: **Input:** RTL design, $\mathscr{P} = \{Plaintext\}, \{Key_0, Key_1\}$
3: **Output:** $Set_{Vulnerable}$
4: **for** $Key_j \in \{Key_0, Key_1\}$ **do**
5: **for** $Plaintext_i \in \mathscr{P}$ **do**
6: $SAIF_i^{Key_j} \leftarrow VCS(Plaintext_i, Key_j)$
7: **end for**
8: **end for**
9: **for all** $Block_i \in \mathscr{M}$ **do**
10: $TC^0_{Block_i} \leftarrow \{SAIF_1^{Key_0}, \ldots, SAIF_n^{Key_0}\}$
11: $TC^1_{Block_i} \leftarrow \{SAIF_1^{Key_1}, \ldots, SAIF_n^{Key_1}\}$
12: $f_i^0 \leftarrow \mathcal{N}(\mu_{TC^0_{block_i}}, \sigma^2_{TC^0_{block_i}})$
13: $f_i^1 \leftarrow \mathcal{N}(\mu_{TC^1_{block_i}}, \sigma^2_{TC^1_{block_i}})$
14: $KL_i \leftarrow D_{KL}(f_i^0, f_i^1)$
15: $SR_i \leftarrow ML(f_i^0, f_i^1, n\ Plaintexts)$
16: **if** $SR_i > SR_{threshold}$ or $KL_i > KL_{norm.th}$ **then**
17: $Set_{Vulnerable} \leftarrow Block_i$
18: **end if**
19: **end for**
20: **end procedure**

III. Evaluation Metrics

In order to reduce time-to-market and the overall cost of adding security to the design, the RTL power side-channel vulnerability evaluation metrics are proposed to perform a design-time evaluation of PSC vulnerability. *To be specific, Kullback-Leibler (KL) divergence metric and success rate (SR) metric based on maximum likelihood estimation are developed and combined to evaluate vulnerability of a hardware implementation.* The first evaluation metric, i.e., KL divergence metric, estimates statistical distance between two different probability distributions, which is defined as follows [13]:

$$D_{KL}(k_i||k_j) = \int f_{T|k_i}(t) \log \frac{f_{T|k_i}(t)}{f_{T|k_j}(t)} dt \qquad (1)$$

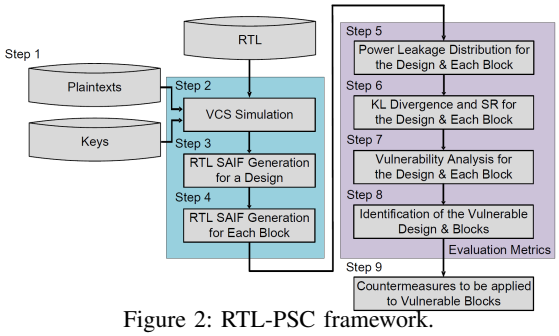


Figure 2: RTL-PSC framework.

He, Miao (Tony); Park, Jungmin; Nahiyan, Adib; Vassilev, Apostol; Jin, Yier; Tehranipoor, Mark.
"RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level."
Paper presented at 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, United States. April 23, 2019 - April 25, 2019.

where $f_{T|k_i}(t)$ and $f_{T|k_j}(t)$ are the probability density functions of the switching activity given keys k_i and k_j, respectively.

For instance, if power leakage probability distributions based on two different keys are distinguishable, KL divergence between these two distributions is high, which provides indication on how vulnerable the implementation is. Hence, KL divergence is suitable for the vulnerability comparison between different implementations. However, the KL divergence value may be difficult to interpret when performing vulnerability analysis for just one implementation. To address this issue, we introduce the second evaluation metric, named success rate (SR)[‡] metric based on maximum likelihood estimation. The SR value represents the probability to reveal the correct key and is suitable to evaluate vulnerability for only one implementation. We derive the SR metric as follows: we assume that the probability density function of the switching activity T given a key K follows a Gaussian distribution, which can be expressed as follows:

$$f_{T|K}(t) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(t-\mu_k)^2}{2\sigma_k^2}} \tag{2}$$

where μ_k and σ_k^2 are the mean and variance of T, respectively. The likelihood function is defined as $\mathscr{L}(k;t) = \frac{1}{n}\sum_{i=1}^{n} \ln f_{T|K}(t_i)$. Based on the maximum likelihood estimation, an adversary typically selects a guess key $\hat{k}$ as follows:

$$\hat{k} = \arg\max_{k\in K} \mathscr{L}(k;t) = \arg\max_{k\in K} \frac{1}{n}\sum_{i=1}^{n} \ln f_{T|K}(t_i) \tag{3}$$

If the guess key ($k_g = \hat{k}$) is equal to the correct key (k^*), the side-channel attack is successful. Thus, the success rate can be defined as follows:

$$SR = \Pr[k_g = k^*] = \Pr[\mathscr{L}(k^*;t) > \mathscr{L}(\langle\bar{k^*}\rangle;t)] \tag{4}$$

where $\langle\bar{k^*}\rangle$ denotes all wrong keys, i.e., the correct key k^* is excluded from $\{k_1, k_2, \ldots, k_{n_k-1}\}$.

KL divergence is closely related to SR since the mathematical expectation of $\mathscr{L}(k^*;t) - \mathscr{L}(k_i;t)$ in Equation (4) is equal to KL divergence between $T|k^*$ and $T|k_i$ [14]. Hence, SR increases accordingly as KL divergence increases. Due to the relation between KL divergence and SR based on maximum likelihood estimation, the combination of KL and SR is proposed for leakage assessment.

A. Selection of a Key Pair

As shown in Algorithm 1, first, a key pair is specified, then the probability distributions of the switching activity based on that key pair can be estimated using Equation (2). The best key pair among all possible pairs is expected to provide the maximum KL divergence for vulnerability evaluation in the worst-case scenario. However, it is impossible and impractical to find the best key pair since the key space is huge, i.e., $\binom{2^{128}}{2}$. Alternatively, an appropriate key pair is able to be chosen, which satisfies the following conditions:

1) Assuming that each set of plaintexts is randomly generated, each key consists of the same subkey, e.g., $Key_0 = \{subkey \ldots subkey\} = 0x151515\ldots15$.
2) Hamming distance (HD) between two different subkeys is maximum, i.e., $subkey$ and $\overline{subkey}$.
3) If $D_{KL}(Key_0||Key_i)$ increases asymptotically as i increases, $i = 1,\ldots n$, Key_0 and Key_n are the appropriate key pair, where Key_i is defined as $[subkey \ldots \underbrace{subkey\ \overline{subkey}}_{\text{i times}}]$.

[‡]This SR is the theoretical SR with infinite plaintexts and SR_{em} in Section IV-C represents the empirical SR based on actual SCA attacks with n plaintexts.

These key pairs can be used for the post-silicon validation. While it is difficult to measure the isolated leakage of each block by a random key pair at the post-silicon stage, the leakage of each block can be measured at the pre-silicon stage through applying the key pairs satisfying the above conditions. Moreover, the evaluation metrics would create the worst-case scenario through applying the key pairs with the maximum Hamming distance.

The selected keys applied to an AES RTL design are 16 pairs of keys starting from all $0s$ key until all Fs key. Each key has 128-bit and Hamming-distance between Key_i and Key_{i+1} is eight, which is shown in Table I. Also, to take into account not only the key's impact on the power consumption, but also the plaintext's impact on power consumption, we use one thousand random plaintexts with the selected key pairs. We use the AES cipher operation itself for the generation of pseudo random plaintext [15]. We use $Plaintext_0$ as seed and then use each ciphertext (j) as the next plaintext ($j+1$),[§]

$$\begin{aligned} &Plaintext_0 = 000...000 \\ &Plaintext_{j+1} = AES(Key_i, Plaintext_j), j = 0,1,...,999. \end{aligned} \tag{5}$$

It can be noted that the key pair Key_0 and Key_{16} in Table I satisfies the above conditions. Furthermore, it can be seen that Key_0 would create a state similar as the reset state of the design, hence, Key_0 and Key_{16} would create the worst case scenario. Therefore, this key pair is used for both the evaluation and the validation metrics, which is shown in Section IV.

Table I: Keys used in RTL-PSC framework.

Key_0	0x0000_0000_0000_0000_0000_0000_0000_0000
Key_1	0x0000_0000_0000_0000_0000_0000_0000_00FF
	.....
Key_{15}	0x00FF_FFFF_FFFF_FFFF_FFFF_FFFF_FFFF_FFFF
Key_{16}	0xFFFF_FFFF_FFFF_FFFF_FFFF_FFFF_FFFF_FFFF

B. Identification of Vulnerable Designs and Blocks

When the appropriate key pair is applied to a design, the same subkey patterns will be propagated to the same blocks, e.g., Sbox blocks in the AES design. The switching activity of each block and the entire design is recorded into SAIF files using VCS functional simulation, which corresponds to power leakage at RTL. The higher the difference between two power leakage distributions is, the higher impact the key has on the power consumption of the design/blocks, the more susceptible the design/blocks are to power analysis attack. Using the KL divergence and SR metrics, the vulnerable design and blocks within the design can be identified. If KL divergence or SR of any design or block is greater than $KL_{threshold}$ or $SR_{threshold}$, the design or the block is considered to be the vulnerable one (Line 16 in Algorithm 1). $SR_{threshold}$ is determined based on the security constraint, e.g., 95% SR with n plaintexts, while $KL_{threshold}$ is determined based on the $SR_{threshold}$ through the relation between KL divergence and SR.

IV. Results and Analysis

In this section, we perform side-channel vulnerability assessment of two different implementations of AES algorithm using RTL-PSC. First, we provide a brief description of the two AES designs: AES Galois Field (GF) and AES Lookup Table (LUT). Then we present the evaluation results generated by RTL-PSC. We validate the accuracy of RTL-PSC evaluation results using gate-level simulation and FPGA silicon results.

[§]This seed is the same as TVLA's setup [8] and 1000 plaintexts are enough to estimate SCA leakage based on our experiments.

He, Miao (Tony); Park, Jungmin; Nahiyan, Adib; Vassilev, Apostol; Jin, Yier; Tehranipoor, Mark.
"RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level."
Paper presented at 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, United States. April 23, 2019 - April 25, 2019.

A. AES Benchmarks

RTL-PSC framework is applied to AES-GF [16] and AES-LUT [17] encryption designs. Both are open-source designs. In AES-GF design, the AES key expansion and AES round operations occur in parallel. The AES-GF implementation takes 10 clock cycles to encrypt each data block. In contrast, the AES-LUT design first performs the key expansion and stores the expanded key in the key registers. The key expansion takes place once for each key. After the key expansion, the round operation starts and takes 11 clock cycles to encrypt a plaintext, precisely, one clock cycle for XORing a plaintext and the key, and 10 clock cycles for 10 round operations. Furthermore, the AES-GF architecture implements the AES SubByte operation with Galois-field arithmetic, while the AES-LUT architecture implements the AES SubByte operation with a lookup table.

The hierarchy of AES-GF and AES-LUT designs is as follows. The AES-GF consists of five SubByte blocks and four MixColumn blocks. Each SubByte block includes four Sbox blocks, each of which includes a GFinvComp block. On the other hand, the AES-LUT cipher consists of a SubWord block, a SubByte block and four MixColumn blocks. SubWord block performs the key expansion operation and therefore, is not related with the plaintext and it is not considered as a potential vulnerable block in the following analysis.

B. Analyzing Suitability of the Key Pairs

We first analyze that the key pairs selected for RTL-PSC evaluation adheres the conditions described in Section III-A. Figures 3(a) and 3(b) illustrate the leakage assessment results at RTL. The switching activities of all blocks during 11 clock cycles are calculated. In the AES-GF implementation, the KL divergence of the design (i.e., AES_GF_ENC) and each block increases asymptotically with increasing the Hamming distance of the key pairs as shown in Table I. Similarly, in the AES-LUT implementation, the KL divergence of the design (i.e., AES_LUT_ENC) and each block also increases asymptotically as the Hamming distance of the key pairs increases. It can be observed that the specified key pair ($key_0 = 0x00\ldots00, key_{16} = 0xFF\ldots FF$) satisfies the key selection conditions, hence is appropriate for RTL side-channel leakage assessment.

C. RTL Evaluation Metrics

In order to determine if a RTL design is vulnerable, KL divergence of the design per clock cycle is calculated. Figure 4(a) shows KL divergence from the second clock cycle to the 11th clock cycle of both AES-GF and AES-LUT RTL implementations, during which 10 round operations for an encryption are performed. At the second clock cycle corresponding to the first round operation, which is mostly exploited by power analysis attack, KL divergence of AES-GF and AES-LUT implementations are 0.47 and 0.28, respectively. These values correspond to 95% SR_{em}[¶] vulnerability level ($SR_{threshold}$) with 25 *plaintexts* and 35 *plaintexts*, respectively, as shown in Figure 4(b). Based on KL divergence metric, as shown in Figure 4(a), KL divergence of AES-GF implementation at the second clock cycle (i.e., 0.47) is greater than that of AES-LUT implementation (i.e., 0.28). Likewise, based on SR metric, as shown in Figure 4(b), the DPA attack success rate of AES-GF implementation is higher than that of AES-LUT implementation with the same number of plaintexts. Similarly, the number of plaintexts required for successful DPA attack on AES-GF implementation is less than that on AES-LUT implementation. Hence, compared with AES-LUT implementation, AES-GF implementation is identified as the more vulnerable design.

Vulnerable Block Identification: Once a RTL deign is determined as a vulnerable one, the next step is to identify the vulnerable blocks within the design that contribute to side-channel leakage significantly. Side-channel vulnerability can be evaluated in both time and spatial/modular domains. In other words, the evaluation metrics based on the switching activity of each block within the design are calculated at fine-granularity scale so that vulnerable blocks can be identified per clock cycle.

First, KL divergence in both time and spatial/modular domains is normalized, i.e., KL divergence of each block is divided by the maximum KL divergence of those blocks ($KL_{norm.th} = KL_i/max(KL_i)$). Then, if the normalized KL divergence of any block is greater than $KL_{norm.th} = 0.5$, that block is identified as the vulnerable one and included into the set of vulnerable blocks. Figure 5 shows KL divergence of each block in both time and spatial/modular domains. The identified vulnerable blocks (i.e., KL divergence greater than $KL_{norm.th} = 0.5$) are denoted with blue bars. Specifically, in the AES-GF design, GFinvComp blocks within Sbox0 and Sbox1 blocks are identified as the vulnerable ones; in the AES-LUT design, SubByte blocks are identified as the vulnerable ones. It should be noted that the threshold values ($KL_{threshold}$ and $KL_{norm.th}$) can be adjusted by the SR vulnerability level.

Evaluation Time: The evaluation time of RTL-PSC includes VCS functional simulation time of the RTL design as well as the data processing time required for analyzing the SAIF files. The evaluation time of RTL-PSC for AES-GF is 46.3 minutes and for AES-LUT is 24.03 minutes. If the same experiments were performed at gate-level designs, it would take around 31 hours. Therefore, our RTL-PSC is almost 42X more efficient as compared to similar gate-level assessment. The evaluation time at layout level is going to be even more expensive (more than a month). RTL-PSC also provides the flexibility to make design changes whereas, the post-silicon assessment provides no flexibility. In the next subsection, we will validate that RTL-PSC is not only efficient, but also accurate using the post-silicon results.

D. Validation of RTL-PSC

The RTL-PSC results are validated through both gate-level and FPGA implementations. We present that the PSC vulnerability assessment results generated by RTL-PSC is highly correlated to the assessment results retrieved from gate-level and FPGA.

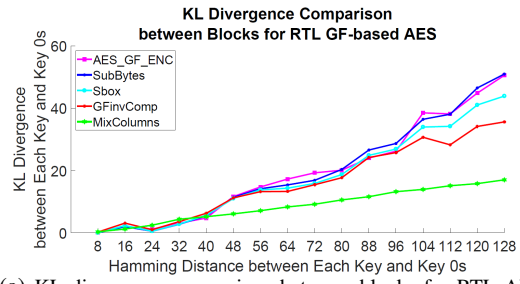
GTL Validation: For gate-level validation, we first synthesize the RTL codes of AES-GF and AES-LUT to gate-level netlist using Synopsys Design Compiler [18] with Synopsys standard cell library. Next we utilize VCS to perform functional simulation of the netlist with the same plaintexts and keys as used in RTL, and generate the corresponding SAIF files. Then, we use Synopsys PrimeTime [18] as well as the generated SAIF files to report the power consumption for the entire design and each block inside the design. Finally, we derive the gate-level KL divergence metric for the design and each block using Equation 1.

We then calculate the Pearson correlation coefficient between the KL divergence of the RTL and GTL design/blocks (shown in Column 2 of Table III and Table II). The high correlation coefficient values ($> 90\%$) indicate that the PSC vulnerability assessment results generated by RTL-PSC at RTL is almost as accurate as the assessment results retrieved from GTL.

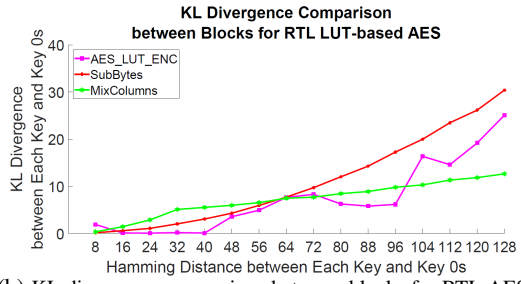
Next we validate the vulnerable block identification results produced by RTL-PSC. Table II presents the correlation coefficient values between the KL divergence metric for each block at RTL and gate-level. The KL metric for each block of

[¶]The SR_{em} represents the empirical SR based on actual SCA attacks with n plaintexts.

He, Miao (Tony); Park, Jungmin; Nahiyan, Adib; Vassilev, Apostol; Jin, Yier; Tehranipoor, Mark.
"RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level."
Paper presented at 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, United States. April 23, 2019 - April 25, 2019.

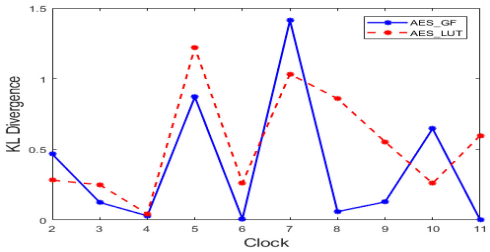


(a) KL divergence comparison between blocks for RTL AES-GF implementation

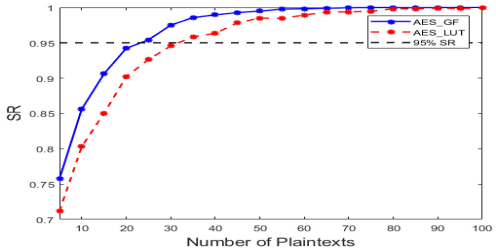


(b) KL divergence comparison between blocks for RTL AES-LUT implementation

Figure 3: KL divergence comparison between blocks for RTL AES-GF and AES-LUT implementations.

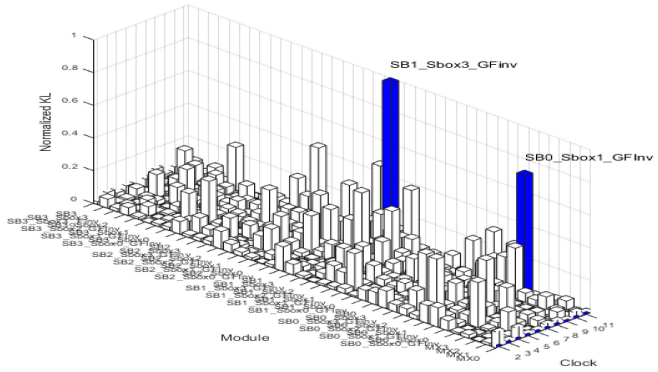


(a) KL divergence per clock cycle for AES-GF and AES-LUT implementations

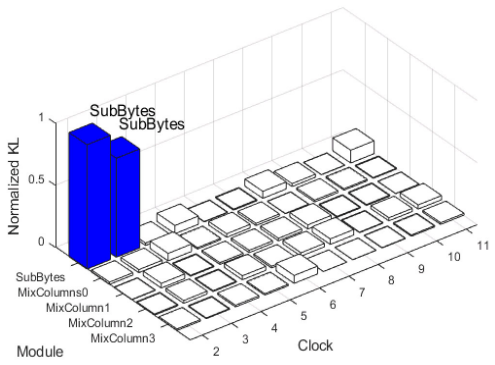


(b) SR_{em} corresponding to KL divergence (0.47 and 0.28) for AES-GF and AES-LUT implementations

Figure 4: KL divergence and SRs for AES-GF and AES-LUT implementations.



(a) Normalized KL divergence for AES-GF implementation in both time and spatial/modular domains



(b) Normalized KL divergence for AES-LUT implementation in both time and spatial/modular domains

Figure 5: Normalized KL divergence for vulnerable blocks within AES-GF and AES-LUT implementations ($KL_{norm.th} = 0.5$).

AES-GF and AES-LUT implementations at RTL has a high correlation to that at gate-level indicating that our vulnerable block identification technique at RTL is as accurate as gate-level. Next we validate the RTL-PSC evaluation results w.r.t. FPGA.

FPGA Validation: For FPGA silicon validation, we use the SAKURA-G board [19] for AES implementations, which contains two SPARTAN-6 FPGAs and is designed for research and development on hardware security. Tektronix MDO3102 oscilloscope is used to measure the voltage drop between shunt registers connected to the Vdd pin. The clock frequency of the AES implementation is 24 MHz. The sampling rate and bandwidth of the oscilloscope are 500 MS/s and 250 MHz, respectively. Figure 6 shows the experimental setup for the FPGA validation.

We first map the AES-GF and AES-LUT designs on an FPGA and then, apply the same plaintexts and keys as used at RTL, and measure the power consumption during encryption operation. Following this, we derive the KL divergence metric from the collected power traces. Then, we can calculate the Pearson correlation coefficient between the KL divergence at RTL and FPGA (as shown in Column 3 of Table III). The high correlation coefficient values ($> 80\%$) indicate that the PSC vulnerability assessment results generated by RTL-PSC at RTL is almost as accurate as FPGA assessment. In other words, RTL-PSC can accurately analyze PSC vulnerability at RTL.

Note that many implementation details, e.g., glitches caused by the gate delay, clock gating, datapath gating, retiming, clock and power network structure is not available at RTL unlike gate-level and FPGA implementations. In spite of these limitations, RTL-PSC is able to accurately identify PSC

He, Miao (Tony); Park, Jungmin; Nahiyan, Adib; Vassilev, Apostol; Jin, Yier; Tehranipoor, Mark.
"RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level."
Paper presented at 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, United States. April 23, 2019 - April 25, 2019.

Table II: Correlation coefficient between KL divergence of RTL and GTL blocks.

AES-GF Blocks, RTL vs. GTL				AES-LUT Blocks, RTL vs. GTL	
SubByte	Sbox	GFinvComp	MixColumn	SubByte	MixColumn
99.11%	99.55%	99.64%	94.73%	99.71%	96.80%

Table III: Correlation coefficient between KL divergence at RTL, GTL, and FPGA silicon level.

Benchmark	RTL vs. GTL	RTL vs. FPGA Silicon Level
AES-GF	99.57%	98.83%
AES-LUT	90.35%	80.80%

vulnerability which proves the efficacy of the framework.

Also, note that it is not feasible to perform vulnerable block identification at FPGA. The reason is that, it is not feasible to isolate the power traces associated to each block from the post-silicon power measurements. Therefore, we could not validate vulnerable block identification at FPGA.

Comparison with State-of-the-art: Veshchikov *et al.* [20] presented a comprehensive survey of simulators for side-channel analysis, e.g., PINPAS [21], SCARD [22], OSCAR [23], etc. These simulators mostly support software crypto algorithms implemented in microprocessors, not hardware crypto modules. On the other hand, TVLA [8] and χ^2-test [12] can work with hardware crypto designs. However, these techniques only provide a pass/fail test and do not provide the quantitative amount of leakage. AMASIVE framework [10] identifies the hypothesis function for HW/HD model to be used for side-channel vulnerability assessment. The major limitation is that it can only identify the hypothesis function and the final vulnerability assessment still needs to be carried out on a prototype device. In contrast, RTL-PSC can quantitatively and accurately assess PSC leakage of hardware crypto modules in RTL level in time efficient manner.

V. Conclusion and Future Work

In this paper, we proposed an automatic evaluation framework to perform a design-time evaluation of side-channel attacks resistance for a cryptographic implementation at RTL. Instead of measuring dynamic power, switching activity at RTL is exploited by using VCS functional simulation to estimate power profile to ensure the evaluation framework is efficient, effective, and library independent. Once the power estimation is complete, the KL divergence metric and SR metric based on maximum likelihood estimation are combined to identify the vulnerable design and blocks within the design. Experimental results on AES-GF and AES-LUT implementations demonstrated that the methodology proposed in this paper were able to perform leakage assessment and identify the vulnerable design/blocks efficiently, effectively, and precisely.

In our future work, the proposed evaluation framework will be applied to evaluate vulnerabilities of the DPA protected designs, thus providing information in terms of scaling and effectiveness for leakage assessment of protected ones. RTL-PSC will be an important component of the recent academic and industrial initiatives for developing CAD frameworks which aim at automating the security vulnerability assessment of hardware designs at design stages [24–28].

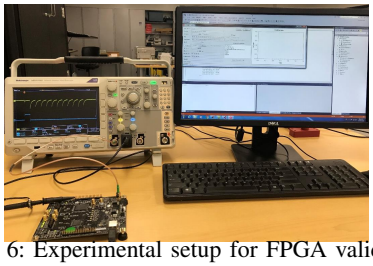


Figure 6: Experimental setup for FPGA validation.

References

[1] P. C. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, ser. CRYPTO '99, 1999, pp. 388–397.

[2] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Cryptographic Hardware and Embedded Systems - CHES 2004*, M. Joye and J.-J. Quisquater, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 16–29.

[3] S. Chari, J. R. Rao, and P. Rohatgi, "Template attacks," in *Cryptographic Hardware and Embedded Systems - CHES 2002*, B. S. Kaliski, ç. K. Koç, and C. Paar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 13–28.

[4] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel, "Mutual information analysis," in *Cryptographic Hardware and Embedded Systems – CHES 2008*, E. Oswald and P. Rohatgi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 426–442.

[5] T.-H. Le, J. Clédière, C. Canovas, B. Robisson, C. Servière, and J.-L. Lacoume, "A proposition for correlation power analysis enhancement," in *CHES 2006*, L. Goubin and M. Matsui, Eds., 2006.

[6] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards (Advances in Information Security)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[7] T. S. Messerges, E. A. Dabbish, and R. H. Sloan, "Examining smart-card security under the threat of power analysis attacks," *IEEE Trans. Comput.*, vol. 51, no. 5, pp. 541–552, May 2002.

[8] G. Becker, J. Cooper, E. DeMulder, G. Goodwill, J. Jaffe, G. Kenworthy, T. Kouzminov, A. Leiserson, M. Marson, P. Rohatgi *et al.*, "Test vector leakage assessment (TVLA) methodology in practice," in *International Cryptographic Module Conference*, vol. 1001, 2013, p. 13.

[9] B. Gierlichs, K. Lemke-Rust, and C. Paar, "Templates vs. stochastic methods," in *Cryptographic Hardware and Embedded Systems - CHES 2006*, L. Goubin and M. Matsui, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 15–29.

[10] S. A. Huss, M. Stöttinger, and M. Zohner, "Amasive: an adaptable and modular autonomous side-channel vulnerability evaluation framework," in *Number Theory and Cryptography*. Springer, 2013, pp. 151–165.

[11] S. A. Huss and O. Stein, "A novel design flow for a security-driven synthesis of side-channel hardened cryptographic modules," vol. 7, no. 1. Multidisciplinary Digital Publishing Institute, 2017, p. 4.

[12] A. Moradi, B. Richter, T. Schneider, and F.-X. Standaert, "Leakage detection with the x2-test," *IACR Transactions on Cryptographic Hardware and Embedded Systems*, vol. 2018, no. 1, pp. 209–237, 2018.

[13] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, no. 1, pp. 79–86, 03.

[14] Y. Fei, A. A. Ding, J. Lao, and L. Zhang, "A statistics-based fundamental model for side-channel attack analysis," Cryptology ePrint Archive, Report 2014/152, 2014.

[15] S. S. Keller, "Nist-recommended random number generator based on ansi x9. 31 appendix a. 2.4 using the 3-key triple des and aes algorithms," *NIST Information Technology Laboratory-Computer Security Division, National Institute of Standards and Technology*, 2005.

[16] A. Lab., "Galois field based aes verilog design," 2007. [Online]. Available: http://www.aoki.ecei.tohoku.ac.jp/crypto

[17] S. Lab., "Lookup table based aes verilog design," 2017. [Online]. Available: http://satoh.cs.uec.ac.jp/SAKURA/hardware/SAKURA-G.html

[18] "Synopsys," http://www.synopsys.com/, 2018, accessed: 2018-04-20.

[19] S. Lab., "Sakura-g fpga board," 2013. [Online]. Available: http://satoh.cs.uec.ac.jp/SAKURA/hardware/SAKURA-G.html

[20] N. Veshchikov and S. Guilley, "Use of simulators for side-channel analysis," in *Security and Privacy Workshops (EuroS&PW), 2017 IEEE European Symposium on*. IEEE, 2017, pp. 104–112.

[21] J. den Hartog, J. Verschuren, E. de Vink, J. de Vos, and W. Wiersma, "Pinpas: a tool for power analysis of smartcards," in *IFIP International Information Security Conference*. Springer, 2003, pp. 453–457.

[22] M. Aigner, S. Mangard, F. Menichelli, R. Menicocci, M. Olivieri, T. Popp, G. Scotti, and A. Trifiletti, "Side channel analysis resistant design flow," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*. IEEE, 2006, pp. 4–pp.

[23] C. Thuillet, P. Andouard, and O. Ly, "A smart card power analysis simulator," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 2. IEEE, 2009, pp. 847–852.

[24] K. Xiao, A. Nahiyan, and M. Tehranipoor, "Security rule checking in ic design," *Computer*, vol. 49, no. 8, pp. 54–61, 2016.

[25] A. Nahiyan, K. Xiao, D. Forte, and M. Tehranipoor, "Security rule check," in *Hardware IP Security and Trust*. Springer, 2017, pp. 17–36.

[26] A. Nahiyan, K. Xiao, K. Yang, Y. Jin, D. Forte, and M. Tehranipoor, "Avfsm: a framework for identifying and mitigating vulnerabilities in fsms," in *Proceedings of the 53rd Annual Design Automation Conference*. ACM, 2016, p. 89.

[27] A. Nahiyan, M. Sadi, R. Vittal, G. Contreras, D. Forte, and M. Tehranipoor, "Hardware trojan detection through information flow security verification," in *Test Conference (ITC), 2017 IEEE International*. IEEE, 2017, pp. 1–10.

[28] A. Nahiyan, F. Farahmandi, P. Mishra, D. Forte, and M. Tehranipoor, "Security-aware fsm design flow for identifying and mitigating vulnerabilities to fault attacks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2018.

He, Miao (Tony); Park, Jungmin; Nahiyan, Adib; Vassilev, Apostol; Jin, Yier; Tehranipoor, Mark.
"RTL-PSC: Automated Power Side-Channel Leakage Assessment at Register-Transfer Level."
Paper presented at 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, United States. April 23, 2019 - April 25, 2019.

Soft MUD

Implementing Manufacturer Usage Descriptions on OpenFlow SDN Switches

Mudumbai Ranganathan
Doug Montgomery
Omar El Mimouni
Advanced Networking Technologies Division
National Institute of Standards and Technology
Gaithersburg, Maryland, USA
e-mail: mranga@nist.gov, dougm@nist.gov, omarilias.elmimouni@nist.gov

***Abstract* -- A Manufacturer Usage Description (MUD) is a generalized network Access Control List that allows manufacturers to declare intended communication patterns for devices. Such devices are restricted to only communicate in the manner intended by the manufacturer, thus reducing their potential to launch Distributed Denial of Service attacks. We present a scalable implementation of the MUD standard on OpenFlow-enabled Software Defined Networking switches**.

Keywords- IOT; MUD; Network Access Control.

I. INTRODUCTION

Internet of Things (IOT) devices (henceforth called "devices") are special purpose devices that have dedicated functions. Such devices typically have communication requirements that are known to the device manufacturer. For example, printer might have the following requirement: Allow access for the printer (LPT) port, local access on port 80 (HTTP) and deny all other access. Thus, anyone can print to the printer, but local access would be required for the management interface which runs on port 80 as a web server. All other access would be in violation of the intended use of the device. The idea behind the Manufacturer Usage Description (MUD) [1] is to declare the intended communication pattern to the network infrastructure using a generalized network Access Control List (ACL) which is specified by the manufacturer, the integrator or the deployer of the device. These are realized as network access controls, by which the device can be constrained to the intended communication patterns.

MUD provides an effective defense against malicious agents taking control of the device and subsequently using it to launch attacks against the network infrastructure. It can also prevent compromised devices from attacking other devices on the network. Thus, MUD substantially reduces the threat surface on a device to those communications intended by the manufacturer.

Because the manufacturer cannot know deployment parameters of devices such as device IP addresses and IP addresses of device controllers, MUD defines class abstractions, using which, the MUD ACLs are defined. For example, the manufacturer may state an intent that devices can only communicate with other devices on the local network, or may state an intent that devices may only communicate with other devices made by the same manufacturer, or that devices may communicate with other devices made by a specific manufacturer on a defined port, or that devices may communicate with specific internet hosts or combinations of the behaviors above. To enable such generality, ACLs are defined with placeholders known as *classes*. These place holders are associated with Media Access Control (MAC) or IP addresses when the ACL is deployed on the switch.

In brief, the system works as follows: A device is associated with a MUD URL. The MUD URL is a locator for the MUD ACL file. The MUD server fetches the MUD file for the device from the manufacturer site, verifies its signature and installs network access controls using whatever mechanism the network switches and firewalls provide. There are several mechanisms that may be available for enforcing access control; for example, iptables could be used or the switch may already support an implementation of network ACLs.

In this paper, we describe a scalable design and implementation of the MUD standard on OpenFlow 1.5 [2] capable Software Defined Network (SDN) switches. An OpenFlow switch supports flow rules that are logically arranged in one or more flow tables in the data plane. The switch connects to one or more controllers that can install flow rules in the switch either reactively, when a packet is seen at the controller, or proactively when the switch connects to the controller. Flow rules have a MATCH part and an ACTION part. The MATCH part can match on different parts of the IP and TCP headers. The ACTION part forwards or drops the packet or sends it to the next table. As packets hit flow rules, metadata can be associated with the packet to provide a limited amount of state as packet processing proceeds from one table to the next. More details are found in [2].

In related work, Hamza et al. [3] consider how MUD may be used with real-world devices to build an Intrusion

Ranganathan, Mudumbai.
"Soft MUD: Implementing Manufacturer Usage Descriptions on OpenFlow SDN Switches."
Paper presented at The Eighteenth International Conference on Networks , Valencia, Spain. March 25, 2019 - March 29, 2019.

Detection System (IDS). They present a simulation based on captured trace data. Details on how to organize flow tables to implement MUD are not presented in their work. The focus of our work is different; in our work, we describe how to implement MUD and demonstrate that it can be done in a scalable fashion.

The rest of this paper is organized as follows: In Section II we outline our design; Section III provides an analysis of our design; Section IV describes our implementation; Section V presents emulation results followed by Section VI which gives measurement on a commercially available home/small business router.

Note that certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

II. DESIGN SKETCH

MUD Access Control Entries (ACEs) can be divided into two categories – those that define intent for communication between a device and a named host, and those that define intent for communication between a device and other classes of devices. The former kind presents no scalability challenges and can be easily implemented using MAC and destination IP address match rules. The main challenge with MUD arises when implementing ACEs that define intent for communication between classes of devices or between the device and hosts on the local network.

Figure 1 shows an example "same-manufacturer" ACE. This indicates the intent that the device may communicate with other devices made by the same manufacturer.

```
{
  "name": "myman0-todev",
   "matches": {
          "ietf-mud:mud": {
            "same-manufacturer": [
              null
            ]
          }
        },
        "actions": {
          "forwarding": "accept"
        }
}
```

Figure 1. Example of a "same-manufacturer" ACE.

Similarly, an ACE can be set up that indicates that the device may communicate with other devices on the local network on a specific port. Such ACEs present scalability problems when naively implemented. For example, if the Same Manufacturer ACE were implemented as MAC to MAC flow rules, there can be $O(N^2)$ rules in the flow table (where N is the number of devices belonging to the manufacturer that are associated with the switch). This is unfeasible as an implementation strategy because switches may be limited in ternary content-addressable memory. Similarly, an explosion of rules will result if the Local Networks ACE were implemented in a single table using MAC address to destination IP match flow rules. We seek a solution that is memory scalable and operator friendly. We make the following assumptions:

- Device Identification: Devices are identified using their MAC addresses on the local network and are dynamically associated with MUD URLs at run-time.
- Flexibility: MAC addresses of devices that will be managed at a switch are not known to the network administrator a priori.
- Network Administration: The network administrator configures information about the network - such as the range of local addresses and the controller classes for the Domain Name System (DNS), Network Time Protocol (NTP) and Dynamic Host Configuration Protocol (DHCP) and device controller.

To achieve scalability and flexibility, ACEs are implemented using SDN flow rules in three flow tables. The source and destination MAC address are classified in the first two flow tables and metadata is associated with the packet. The third table implements the MUD ACEs with rules that stated in terms of the packet classification metadata that is assigned in the first two tables.

The flow pipeline is as shown in Figure 2, with the packet being finally sent to a table that implements L2Switch flow rules which is provided by another application.

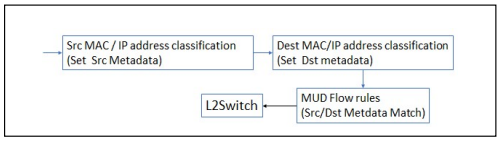


Figure 2. Flow Pipeline structure

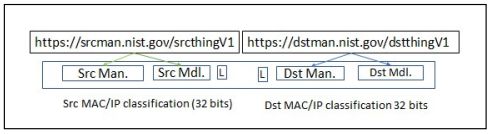


Figure 3. Source and destination metadata assignment.

The OpenFlow metadata field consists of 64 bits. We organize this as two 32-bit segments. Each 32-bit segment encodes a triple <*manufacturer, model, local-networks flag (L)* > as shown in Figure 3. The manufacturer and model are

determined from the MUD URL and the local-network flag is determined from Source / Destination IP address and network configuration.

The packet classification metadata rules are reactively inserted when packets arrive at the switch as follows: When the switch connects to the controller, the source and destination classification tables are initialized with low priority rules that unconditionally send IP packets up to the controller. This generates a *PacketIn* event at the controller which then inserts Source (or destination) MAC match rules that assign metadata to the packet and forward to the next table at a higher priority. Subsequent packets that match on the same MAC will have metadata associated with it and be forwarded to the next table without controller intervention. The next stage is to do the same for the destination MAC match rule.

The controller maintains a table associating MAC address with a MUD URI. This mapping is learned dynamically during DHCP processing, i.e., when the device sends out its own MUD URL when requesting an address (using the newly defined DHCP options 161) or it can be configured by the administrator for the device if DHCP support has not been implemented on the device. Each manufacturer (i.e., the "authority" portion of the MUD URL) and model (i.e., the entire MUD URL) is assigned a unique integer, which is placed in the metadata as shown in Figure 3. The controller also has knowledge of what constitutes a "local network" (typically the local subnet) which is assigned a bit in the metadata. If a MAC address does not have a MUD URL associated with it (e.g., a laptop) then it is assigned an implementation reserved metadata classification of UNCLASSIFIED that cannot be assigned to any real MUD URL.The next table implements the MUD ACEs. Note that at this stage of the pipeline, metadata has already been associated with the packet. MUD rules are implemented as ACCEPT rules. That is, if the metadata assigned to a packet matches the match part of the mud rule, it is sent to the next table.

Default unconditional high priority rules are initially inserted that allow interaction of the device with the reserved ports for DHCP and NTP. These are inserted into the MUD rule table on switch connect with the controller. The DHCP match rule has a "send to controller" Action part so that the controller may extract the MUD URL from the DHCP request if it exists. This enables the controller to associate a MUD URL with the source MAC address based on the DHCP request.

After the MUD URL is associated with the device, the MUD profile is retrieved by the SDN controller and flow rules that implement the MUD ACEs are inserted into the MUD table in the following order:

- High priority source (or destination) metadata and TCP Syn. flag match drop action rule to enforce TCP connection directionality. MUD ACEs can specify which end of a TCP connection is the initiator. For such MUD ACEs, we insert rules that drop packets where the connection is initiated from the wrong direction.
- Lower priority Rules that match on source metadata and destination IP addresses for access to specific named hosts or classes of hosts (e.g. Controller or my-controller) as specified by the MUD ACEs.
- Rules that match on source IP address and destination metadata for inbound packets to the IOT device as specified by the MUD ACEs.
- Rules that match on source and destination metadata for allowing access to manufacturer or model or local network classes as specified by the MUD ACEs.
- Lower priority Drop rule for packets that match on Source Model metadata but do not match on one of the rules above.
- Lower priority Drop rule for packets that match on Destination Model metadata but do not match on one of the higher priority rules above.
- Lower priority default UNCLASSIFIED packet pass through rule. The default MUD behavior allows all packets that are metadata tagged as UNCLASSIFIED to pass through the pipeline.

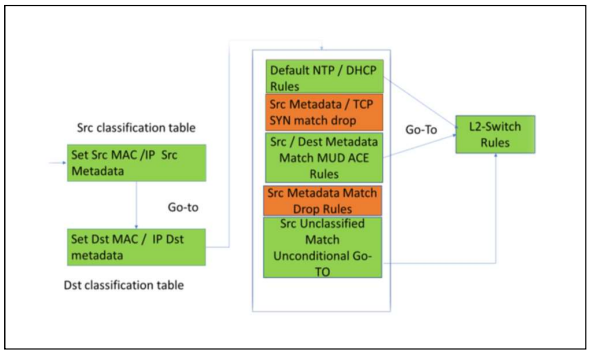


Figure 4. Detailed Flow Pipeline structure. The first two tables are classification tables which assign metadata. The third table is the MUD rules table. Drop rules are color coded Red.

III. ANALYSIS

The scheme we have described above is dynamic and memory scalable with O(N) rules for N distinct MAC addresses at switch. By dividing the rules into packet classification rules and MUD rules which are dependent on the metadata assigned on the first two tables, MUD rules can be installed independently of packet classification. The devices may appear at the switch prior to the MUD rules being installed or vice versa. This allows for dynamic configuration i.e. the MUD ACL table can be changed dynamically at run time without needing to re-configure the rules in the first two tables that classify the packets and vice

versa. The packets may be initially marked as UNCLASSIFIED and later when they are associated with a MUD profile (using the DHCP or other mechanism outlined in the MUD specification), the appropriate metadata is assigned to them.

Because the MUD ACEs are expected to be relatively static and few, the flows in the MUD rule table have hard timeouts to match the cache timeout in the MUD file. This can be in the order of days. The packet classification flow rules have short (configurable) idle timeouts. This limits the size of the table and allows for dynamic adjustment of the table when MAC addresses appear and disappear at the switch. The shorter the idle timeout for the classification rules, the less time it takes for reconfiguration and the less time it takes to purge the table from unreferenced entries. However, the shorter the timeout, the more overhead by way of communication with the controller due to the increased number of *PacketIn* events at the controller. We present experimental results in section V.

Our scheme, as described thus far, requires that a packet must be processed at the controller and a rule installed before packet processing may proceed. The initial rule in the MAC address classification stage that is installed when the switch connects, sends the packet to the controller but not to the next table. Thus, a packet may not proceed in the pipeline before it can be classified. This may be necessary if strict ACL-dictated behavior is required but there are some resultant performance consequences i.e., a disconnected or failed controller causes a switch failure because no packets from a newly arriving device can get through prior to the classification rule being installed.

To address this problem, we loosen up the interpretation of the ACE specification. We define a "relaxed" mode of operation where packets can proceed in the pipeline while classification flow rules are being installed. This may result in a few packets being allowed to proceed, in violation of the MUD ACEs with the condition that the system will become eventually compliant to the MUD ACEs.

To implement this behavior, the initial rule installed in the packet classification table with infinite timeout, allows the packet to proceed through the pipeline and delivers the packet to the controller simultaneously. If the controller is offline or fails during rule installation, the packet is sent to the next table with the initial rule and there is no disruption. When the controller comes online again, it will get a packet notification and install the appropriate rule – thus restoring MUD compliant behavior. Thus, the switch becomes resilient to controller failures, with the failure mode being to allow communication.

However, there is another source of potential disruption that must be addressed: Because the source and destination MAC addresses are classified using two tables, it is possible that the source MAC address classification rule exists in the table, while the destination MAC address classification rule has not yet been inserted into the next table. If the MUD rules have been inserted already, this will result in dropped packets in the MUD rules table until the destination table is populated, because the fall through action for Source MAC - classified packets that do not match a MUD ACE rule is to drop the packet.

We address this issue by defining reserved metadata classifications as follows:

- UNCLASSIFIED: The MAC address does not belong to any known MUD URL. For example, if the packet is emitted with a source address belonging to a laptop, for which no MUD rules exist, then its source MAC address is UNCLASSIFIED.
- UNKNOWN: The MAC address has been sent to the controller and is pending classification. The default rule that is installed when the switch connects to the controller sends the packet to the controller on IP match and stamps the packet with metadata of UNKNOWN.

The classification tables each have rules that send the packet up to the controller while setting the corresponding metadata (for source or destination MAC) to UNKNOWN and forwarding the packet to the next stage. The MUD rules table has rules that permits packets that are UNKNOWN in source or destination MAC classification to proceed to the next stage. The scheme is as shown in Figure 5.

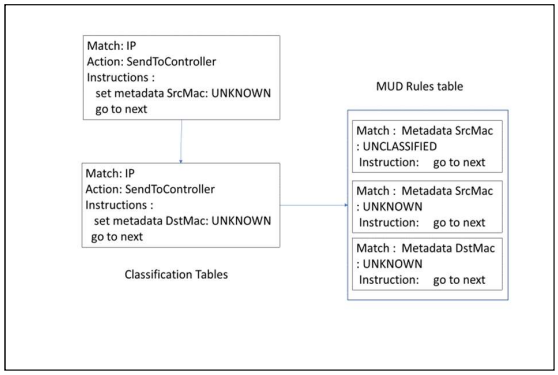


Figure 5. Temporary classifications label added to prevent blocking of flow pipeline during configuration.

On *PacketIn*, the controller pushes flow rules to correctly classify the packet. Because these packet classification rules are pushed at a higher priority than the default send to controller rule in the classification tables, the metadata will change from UNKOWN to the actual classification determined by the controller when the flow rule is installed. In the meanwhile, the pipeline is not blocked.

This "eventually compliant" mode of operation avoids packet drops and provides controller failure resiliency;

however, there some limitations: (1) A few packets that violate the MUD rules could get through prior to the classification rule being installed at the switch. This could result in a temporary violation of the ACEs. (2) TCP direction enforcement for short flows, which depends upon detection of TCP SYN flags and correct classification of MAC addresses, is not possible to enforce at the switch until a flow rule that classifies the packet is installed. We quantify these limitations in the next sections.

IV. Implementation

Our implementation [4] uses the OpenDaylight (ODL) SDN controller [5]. The configuration information for the system, which includes the MUD file and ACLs file are presented as north-bound API, are generated using the ODL YANG tools. The association between MUD URL and MAC can be configured directly or inferred by the controller by examining interactions between IOT devices and the DHCP server. For the performance measurement experiments, we directly configured the MAC to MUD URL association.

V. Emulation Experiments

To measure scalability of the implementation, our experimental scenario on MiniNet [6] consisted of 100 devices on one switch all belonging to the same manufacturer randomly exchanging messages. A device randomly picks another device and sends 10 pings, then sleeps randomly with an exponentially distributed average sleep time of 5 seconds. Our goal is to measure the memory scaling as the idle timeout of flow rules is altered. The following chart shows sum of the maximum number of rules in the source and destination classification tables for different values of the idle timeout.

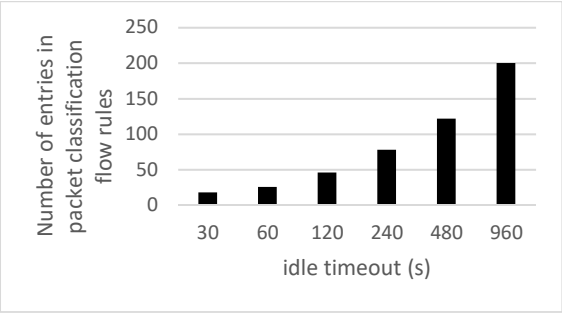


Figure 6. Packet Classification tables size variation with idle timeout. The MUD Rules table is a constant size and has infinite timeout.

In all cases, it is possible to implement the system with just a few rules in the classification table at the expense of an increasing number of *PacketIn* events processed at the controller.

To quantify the overhead involved with *PacketIn* processing under load, we measured the number of packets seen at the controller per burst of packets sent by varying the idle timeout settings for the packet classification rules. The results are shown in Figure 7. The maximum number of *PacketIn* events per burst of pings reaches a maximum of about 6 packets with the classification flow idle timeout set to 15 seconds – 6 packets are processed at the controller under these load conditions before the flow is pushed to the switch. The time it takes for the flow to appear at the switch is the window within which ACE violations can occur in the Relaxed ACL model and is hence significant. Measurements on an actual switch are presented next.

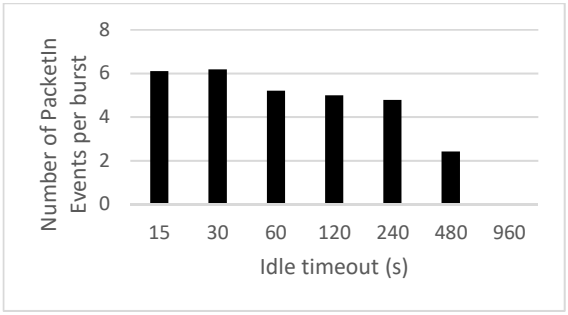


Figure 7. Maximum Number of packetIn events (observed at the controller) per burst of packets.

VI. Measurements on an Omnia Turris Router

To measure how well our implementation would perform on commercially available hardware that may be a part of a home or small business, we tested our implementation an Omnia Turris [7] router that supports OpenVSwitch [8]. Raspberry Pi devices were used for load generation.

We installed a MUD Profile using the DHCP mechanism which allows the device to be accessed on port 80 from any machine on the local network. The device can access www.nist.gov on port 443. All other access is denied. The baseline performance of the router was measured. Relaxed ACLs were used to install the flow rules. Then, using iPerf [9], we measured the bandwidth with the MUD rules installed under different scenarios. This gives an indication of the overhead involved with MUD rule processing. As previously described, relaxed ACLs give us some advantages i.e., resilience to controller failures and reduced latency for packets that do not violate ACLs. However, packets may get through in violation of an ACL until the time a packet classification rule is pushed to the switch and appears in the switch table as a flow. How many packets get through before further communication is blocked? We used iPerf to perform an experiment where the device initiates an outbound connection with a peer on the local network and sends packets to it. As the "attempted bandwidth" is increased, more packets make it through the pipeline before being blocked, reaching a maximum of about 3 MB total leakage before the flow rules are applied as shown in Figure 8.

Ranganathan, Mudumbai.
"Soft MUD: Implementing Manufacturer Usage Descriptions on OpenFlow SDN Switches."
Paper presented at The Eighteenth International Conference on Networks , Valencia, Spain. March 25, 2019 - March 29, 2019.

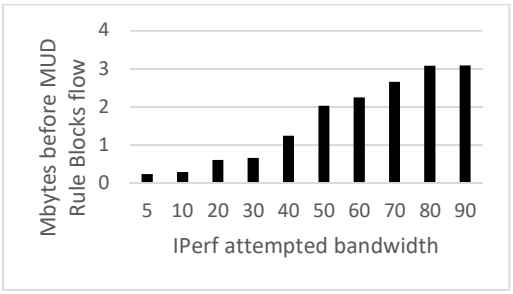


Figure 8. Relaxed ACL TCP packet leakage before ACL application. This is the amount of data that gets through before iperf stops.

Thus, if devices are expected to communicate infrequently and in short bursts, it is better to use the strict ACL model. Otherwise, it is possible that the communication may complete before the MUD ACE flow rules intervene.

Finally, we measured the overhead of strict ACLs on connection establishment. We initiate a connection from a local network resident device to a server accepting connections on port 80 on the MUD-compliant IOT device and measure the overhead in TCP connection establishment with and without relaxed ACL support over 100 attempts. The results are summarized in Table 1.

ACL Model	Max Connection establishment time (s)	Standard deviation (s)
Relaxed	0.002	.0003
Strict	2.0	.15

Table 1: Max TCP connection establishment time for strict and relaxed ACLs.

Significantly worse performance for strict ACL is caused by packets being dropped before flow rules are pushed. Dropping packets when the TCP connection is being established adversely impacts the connection establishment time. Note that this phenomenon only occurs when the rule is first installed because of the round trip to the controller before the installation of the rule.

VII Conclusion

In this paper we presented the design and implementation of the MUD standard on OpenFlow switches, thereby demonstrating its implementation feasibility – even on limited memory devices. Our design is model driven, resilient to controller failure and allows for dynamic re-configuration. Our design uses $O(N)$ flow rules for N distinct MAC addresses seen at the switch.

An open question is how to set the idle timeout for the flows. Our timeout policy for the experiments described in this paper was to set the timeout the same for all devices – which makes sense in this case given a homogenous communication pattern with equal probability that a randomly selected pair of MAC addresses will communicate. In general, the communication between devices and between devices and its controller or host is not likely to be uniform. To achieve best utilization of the switch flow table memory, the idle timeout should be set high for MAC addresses that have a high probability of being referenced and set low for MAC addresses that have a low probability of being referenced [10]. It would be useful to extend the MUD standard to provide hints for communication frequency and length of communication burst for different MUD ACEs so that the controller can use this information to optimize timeouts and pick the appropriate management strategy on the classification flow table.

Our future work includes setting timeouts adaptively and combining MUD with an IDS to develop a comprehensive enterprise security architecture.

References

[1] E. Lear, R. Droms, and D. Romascanu, “Manufacturer Usage Description Specification,” Internet Engineering Task Force Work in Progress, Jun. 2018. https://datatracker.ietf.org/doc/draft-ietf-opsawg-mud/ [Accessed: March 2019]

[2] Open Networking Foundation, “OpenFlow Switch Specification, Version 1.5.1 (Protocol version 0x06)”, https://www.opennetworking.org/software-defined-standards/specifications/ [Accessed: March 2019]

[3] A. Hamza, H. H. Gharakheili, and V. Sivaraman, “Combining MUD Policies with SDN for IoT Intrusion Detection,” in Proceedings of the 2018 Workshop on IoT Security and Privacy, 2018, pp. 1–7.

[4] nist-mud - NIST SDN MUD implementation, https://github.com/usnistgov/nist-mud [Accessed: March 2019]

[5] OpenDaylight, SDN Controller https://www.opendaylight.org [Accessed March 2019 }

[6] MiniNet - An instant Virtual Network On Your Laptop. http://www.mininet.org [Accessed: March 2019]

[7] Omnia Turris https://omnia.turris.cz/ [Accssed: March 2019]

[8] OpenVSwitch: Production Quality Multilayer, Open Virtual Switch., https://openvswitch.org Sep-2018. [Accessed: March, 2019]

[9] J. Dugan, S. Elliott, B. Mah, J. Poskanzer, and K. Prabhu, “iPerf - The ultimate speed test tool for TCP, UDP and SCTP.” https://iperf.fr [Accessed: March 2019]

[10] A. Vishnoi, R. Poddar, V. Mann, and S. Bhattacharya, “Effective switch memory management in OpenFlow networks,” Proc. 8th ACM Int. Conf. Distributed Event-Based Systems, 2014, pp. 177-188.

Electric Field Gradient Reference Material for Scanning Probe Microscopy

J. J. Kopanski, M. Fu[1], and L. You

National Institute of Standards and Technology
Nanoscale Device Characterization Division
Gaithersburg, MD 20899 USA

INTRODUCTION

Electrical scanning probe microscopes (eSPMs), such as the scanning Kelvin force microscope (SKFM), scanning capacitance microscope (SCM), or various scanning microwave microscopes (SMMs) are sensitive to the electric field between the sample and tip. Interpretation of measurements with these techniques can be confounded due to unknown tip shape and volume of interaction with the sample. Any two-terminal electrical measurement of electric field, capacitance, resistance, or inductance depends on the shape of the electrodes at each terminal. For simple two-dimensional metal-insulator capacitors, $C = \varepsilon_0\varepsilon_i A/t_i$ where C is capacitance, ε_o is the electric constant, ε_i the relative permittivity of the insulator, A is the device area, and t_i is the insulator thickness. For two-dimensional resistors, $R = \rho L/A$, where R is the resistance, ρ the material resistivity, and L the device length. Without knowing the device geometries, the material electrical properties (in these examples, ε_i and ρ) cannot be deduced regardless of how accurately the capacitance or resistance is measured. For complex electrode shapes varying in three dimensions, such as eSPMs, the accuracy of extracted material properties depends on having detailed information about the shape of the electrodes.

Any eSPM measurement of a spatially varying electric field at the surface of a sample has a large uncertainty due to the unknown details of the tip shape near the surface. We have designed an electric field gradient reference sample to provide an unambiguous known reference sample of transitions in electric field over small distances. Since all dimensions and materials of the reference sample are known, the electric field at the surface can be calculated precisely. This will allow the sources of error in the measured signal to be determined as a function of tip shape and measurement conditions. The reference also functions as a method of calibrating SKFM to improve its accuracy and to make calibrated measurements on unknowns.

E-GRAD REFERENCE MATERIAL DESIGN

We designed and built a test structure of 30 interdigitated buried metal lines that when biased will produce a stepped electric field across its length, Figure 1. Each set of three lines is connected to a polysilicon voltage divider that reduces the potential by a factor of 0.382 per stage. If the initial set of lines is biased at 10 V, voltages of 10 V, 3.820 V, 1.459 V, 552 mV, 213 mV, 81.3 mV, 31.1 mV, 11.9 mV, 4.53 mV, 1.73 mV, and 0.66 mV are produced down the structure. The interdigitated lines on the other side of the structure can be grounded, at a backing potential, or 180° out of phase with their opposing interdigitated lines. Structures with four line-to-spacing geometries were produced: 1.2 µm lines with 6.8 µm spacing, 3.2 µm lines with 4.8 µm spacing, 5.2 µm lines with 2.8 µm spacing, and 6.8 µm lines with 1.2 µm spacing. The lines are fabricated beneath an intermetal dielectric of approximately 1 µm thick. The region of the test structure with buried metal is covered by a metal cap that is exposed by the final bonding pad contact cut. This metal cap is removed post chip fabrication by etching.

[1] Mathew Fu contributed to this research as part of the NIST Summer Undergraduate Research Fellowship (SURF) program. He is currently an undergraduate student at Cornell University.

Kopanski, Joseph; You, Lin.
"Electric Field Gradient Reference Material for Scanning Probe Microscopy."
Paper presented at 2019 International Conference on Frontiers of Characterization and Metrology for Nanoelectronics, Monterey, CA, United States. April 2, 2019 - April 4, 2019.

The SKFM typically produces images of the contact potential difference (CPD) between the conducting tip and a metallic surface. Our test structure has buried metal lines covered by a thin insulating layer. In this case, a contact potential cannot be defined as the oxide surface does not have a defined work function like a metal. This test structure effects the SKFM in two ways: through the geometrical capacitance between the metallic lines and metal tip which varies with probe height, z, and through the electrostatic force generated by the potential difference between the biased lines and the SKFM tip. The potential at the oxide surface from a voltage biased buried metal line will depend on the bias voltage, thickness of the insulating cover layer and its dielectric constant. When measured with SKFM this test structure will produce a force on the tip that is the product of the differential capacitance, dC/dz and the applied voltage, with the contact potential replaced by an effective CPD induced by the buried metal lines, Eqn. 1. Setting the amplitude, A, proportional to force F, we can separate the dC/dz component and the bias voltage component of the EFM response, Eqn. 2.

$$F = -\frac{dC}{dz}\left(V_{dc} - V_{CPD,eff}\right)V_{ac} \quad [1]$$

$$V_{CPD,eff} = \frac{A}{\left[-\frac{dC}{dz}V_{ac}G(\omega)\right]} + V_{dc} \quad [2]$$

Where $G(\omega)$ is the transfer function relating the force, F, and the resulting amplitude of oscillation, A. Detailed three dimensional simulations of the test structure are underway using COMSOL[2] MultiPhysics. Simulation will yield the expected electric field at the test structure surface and the potential difference measured with various tip shapes.

REMOTE BIAS ELECTROSTATIC FORCE MICROSCOPY

Biased, subsurface structures can be imaged with SKFM or electrostatic force microscopy (EFM) [1-2]. We found that the detection depth can be substantially improved, using a technique that we have dubbed remote bias induced EFM (RB-EFM). In RB-EFM, a set of synchronized ac with phase φ plus dc signals are sent to the adjacent sets of buried metal lines, instead of the cantilevered tip. An external ultra-high frequency lock-in amplifier (LIA) (Zürich Instrument, Zürich, Switzerland) was used to provide the excitation signals. To avoid interference with the atomic force microscope (AFM), the excitation near the cantilever second resonance was used. The amplitude and phase of the cantilever oscillation induced by the remote bias was measured from the AFM's position sensitive detector (PSD) signal fed back to the LIA as the input signal, using the frequency of the remote bias as the reference signal. One side of the interdigitated lines are biased in phase, while the other side is biased 180° out of phase. This measurement scheme forces the tip oscillation to flip 180° when scanned from one line to the next with an abrupt transition between the lines. Since for any line the adjacent lines are biased 180° out of phase, the structure itself provides an active shield, allowing the effect of the central line on the cantilever to be isolated.

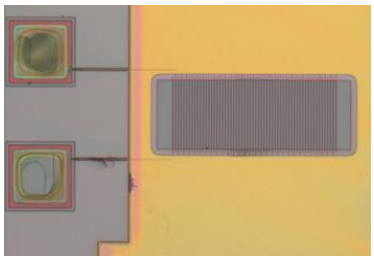
RESULTS

We report initial RB-EFM results here, obtained for a single set of lines biased at a series of discrete voltages (-5 V to +2 V), Figure 2. Each curve is an average of around 10 scans. The measured signal is proportional to the amplitude of cantilever vibration. The signal is forced to a minimum at the mid-point between adjacent lines by the change in phase in the driving signal between adjacent lines. As the dc bias on the lines is increased from -5 V to +2 V, the amplitude steadily decreases. Beyond +2 V, the signal reaches a minimum and begins increasing again with further increases in the dc bias. This is due to the effective contact potential difference induced by the buried metal lines. The RB-EFM response from the lines on the far ends of the test structure do not have a line biased 180° out of phase on their far side, resulting in an enhanced response. Applying Eqn. 2, we can separate the dC/dz component (independent of dc bias), Fig. 3, and the effective CPD from the buried metal structures, Fig. 4. The dC/dz component is due only to the geometry of the metal lines and the SPM tip. The effective contact potential is inherently due to the

[2] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment used are necessarily the best available for the purpose.

Kopanski, Joseph; You, Lin.
"Electric Field Gradient Reference Material for Scanning Probe Microscopy."
Paper presented at 2019 International Conference on Frontiers of Characterization and Metrology for Nanoelectronics, Monterey, CA, United States. April 2, 2019 - April 4, 2019.

materials and any fixed charge of the buried structure. The applied dc bias affects the magnitude of the response. At the center of the line, the induced response scales linearly with the applied dc voltage.



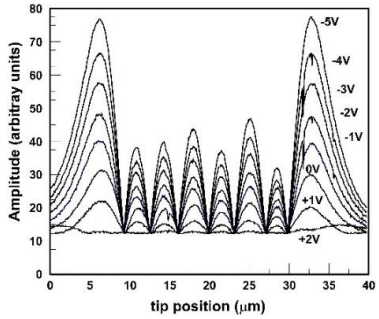
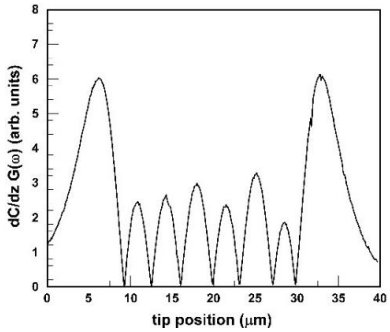


Figure 1: Photo micrograph of the prototype E-Grad Reference Structure. The area surrounding the active area is covered with a ground metal layer, hiding the voltage divider resistors from view.

Figure 2: RB-EFM amplitude measured for a single set of interdigitated lines at dc bias voltages increasing from -5 V to +2 V in 1 V steps.



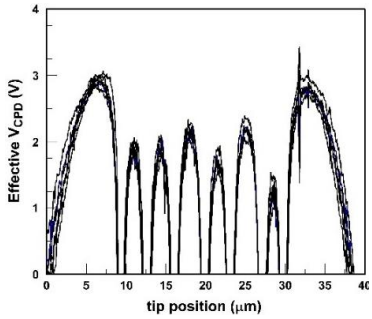


Figure 3: The geometrical component of the response (dC/dz), which depends on the tip to metal line capacitance only. Data is calculated by taking the slope of amplitude versus dc bias voltage at each tip position point from Fig. 2.

Figure 4: Effective contact potential difference at the test structure insulator surface due to the buried metal lines. Eight curves from each dc bias are plotted, approximately the same effective CPD is seen regardless of the dc bias.

CONCLUSIONS AND ONGOING WORK

We built and tested a candidate electric field gradient reference material. Preliminary measurements show that we can separate the geometrical effects due to dC/dz, the effects of the test structure materials and fixed charge as an effective contact potential, and the effects of applied dc bias voltages. Detailed measurements on the complete test structure with a variety of tips are underway. The width of the transition region will provide a measure of spatial resolution. By comparing the measured response for each tip to the expected electric field at the test structure surface, we can extract a figure of merit for a given tip geometry and measurement procedure.

REFERENCES

1. Lin You, Jung-Joon Ahn, Yaw S Obeng and Joseph J Kopanski, J. Phys. D: Appl. Phys. 49 045502, 2016.
2. J. J. Kopanski and L. You, Remote Bias Induced Electrostatic Force Microscopy for Enhanced Subsurface Imaging, in preparation.

KEYWORDS

Electric field, electrical scanning probe microscopy, eSPM, reference materials, standards for SPM.

Kopanski, Joseph; You, Lin.
"Electric Field Gradient Reference Material for Scanning Probe Microscopy."
Paper presented at 2019 International Conference on Frontiers of Characterization and Metrology for Nanoelectronics, Monterey, CA, United States. April 2, 2019 - April 4, 2019.

A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope

Benjamin W. Caplins[a], Jason D. Holm[a], and Robert R. Keller[a]

[a]National Institute of Standards and Technology, 325 Broadway, Boulder, CO, USA 80305

ABSTRACT

Unit cell orientation information is encoded in electron diffraction patterns of crystalline materials. Traditional transmission electron detectors implemented in the scanning electron microscope are highly symmetric and are insensitive to in-plane unit cell orientation information. Herein we detail the implementation of a transmission electron detector that utilizes a digital micromirror array to select anisotropic portions of a diffraction pattern for imaging purposes. We demonstrate that this detector can be used to map the in-plane orientation of grains in two-dimensional materials. The described detector has the potential to replace and/or supplement conventional transmission electron detectors.

Keywords: digital micromirror device (DMD), scanning electron microscope (SEM), scanning transmission electron microscopy (STEM), transmission electron detector

1. INTRODUCTION

The scanning electron microscope (SEM) is widely used for imaging the surfaces of materials over six orders of magnitude in scale (10^{-3} m to 10^{-9} m).[1] In an SEM, a focused electron beam (*ca.* 1 nm diameter on modern field emission SEMs) is scanned across the surface of a bulk sample in a raster pattern. At each point on the sample, the electron beam interacts with the sample and generates a signal that is measured to create an image. By far, the most common signal measured to generate an image is known as the secondary electron (SE) signal, comprised of low-energy inelastic electrons, emitted in the backscatter direction. Generally, these electrons are collected over a large solid angle with an Everhart-Thornley detector[2] and are relatively insensitive to the sample crystallography. When applied to 2D materials, SE images are generally used for their topographic,[1] work function,[3] or material-dependent SE-yield[4] contrast mechanisms.

For nanomaterials in an SEM, where the electron mean free path at 30 keV can be comparable to the material thickness, most of the incident electrons undergo small-angle scattering events and transmit through the sample (Figure 1a). This transmitted electron scattering distribution forms a diffraction pattern in the far-field, and portions of it can be detected to create real-space images. Experimental diffraction patterns of amorphous and crystalline materials are shown in Figure 1b. In the SEM, transmission electron detectors are usually axially symmetric and integrate circular or annular regions of the diffraction pattern (e.g. Figure 1a). When the central disk of the diffraction pattern is spatially integrated, the corresponding image is called a 'bright field' (BF) image. Conversely, if some portion of the area outside the central disc is spatially integrated, the corresponding image is called a 'dark-field' (DF) image. A common type of DF image is the annular dark-field (ADF) image. While these imaging modes enable mass-thickness measurements,[1] they generally do not contain interpretable crystallographic information in an SEM for 2D materials.

To obtain image contrast based on crystallographic orientation of 2D materials, a non-axially symmetric portion of the diffraction pattern needs to be integrated. As an example, we use the six-fold symmetric diffraction

Further author information: (Send correspondence to B.W.C.)
B.W.C.: E-mail: benjamin.caplins@nist.gov
J.D.H.: E-mail: jason.holm@nist.gov
R.R.K.: E-mail: bob.keller@nist.gov

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

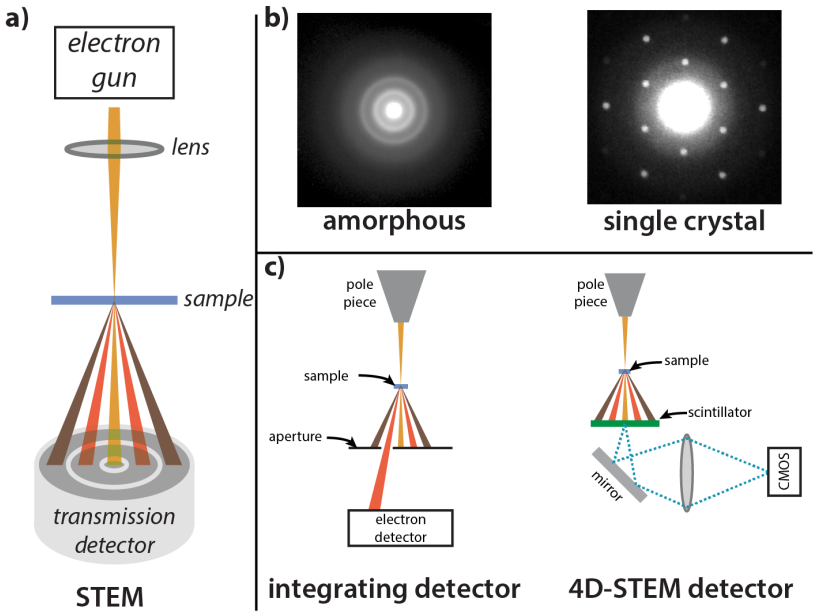


Figure 1. (a) Schematic of a scanning transmission electron microscope. An electron gun emits electrons which are focused onto a thin, electron-transparent sample. The transmitted electrons then strike an electron detector. (b) Example electron scattering distributions (30 keV) incident on a transmission detector. The left image shows an axially symmetric diffraction pattern from an amorphous material, while the right image shows an anisotropic diffraction pattern from a crystalline material. The diffraction pattern on the right is from a single crystal region of monolayer graphene. (c) Two types of detectors capable of discerning anisotropy in crystalline diffraction patterns. A physical mask may be used to isolate a subset of the diffracted electrons, which are then summed on an integrating detector. Alternatively, the full diffraction pattern may be imaged and saved for each beam position resulting in a four dimensional dataset, $I(x, y, k_x, k_y)$ (i.e., 4D-STEM).

pattern of graphene in Figure 1b. To generate orientation contrast, the imaging mask must collect the diffracted electrons from a specific graphene orientation, while rejecting the diffracted electrons from other, rotated domains.

There are currently two methods to accomplish this goal. In the first method, a physical mask is placed over a high-bandwidth integrating detector to select a subset of the scattered electrons (Figure 1c).[5] Alternatively, the active area of the detector is constructed to have a defined shape. This method has the benefit of being simple to implement using high-bandwidth integrating electron detectors that can collect megapixel images in just a few seconds. However, these integrating detectors cannot reconstruct the full diffraction pattern, meaning that it is not always known what scattering conditions are being selected. An additional challenge is that it is difficult to manipulate the orientation of a physical mask/detector inside a vacuum chamber.

In the second method referred to as 4D-STEM, the diffraction pattern is collected for each beam position resulting in a four dimensional dataset, $I(x, y, k_x, k_y)$. Either a pixelated direct electron detector[6] (not commercially available for SEMs at this time) or a scintillator/camera combination[7] (Figure 1c) can be used to collect this data. With the 4D dataset in hand, the in-plane crystallographic orientation can be deduced from the diffraction patterns. Unfortunately, the 4D-STEM method suffers from the relatively slow readout rate of cameras ($\approx 10^3$ frames per second) resulting in $\approx 10^3$ seconds per scan. Furthermore, the large datasets are only sparsely populated with information and are, at present, unwieldy to transfer and analyze.

Herein, a new transmission detector called the programmable scanning transmission electron microscope (p-STEM) detector is described.[8,9] This detector utilizes a digital micromirror device (DMD) to give direct access

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

to the full electron diffraction pattern, with the additional capability to rapidly create targeted DF images, that can emphasize the crystallographic orientation of the sample. In many ways, this detector is a modern realization of a detector suggested by Cowley and Spence in 1979, which used tiny mirrors on actuators in place of a DMD.[10]

As a proof of principle, we use the p-STEM detector to map the in-plane orientation of a graphene sample and generate diffraction contrast from multilayer graphene. To date, orientation mapping of graphene has not been demonstrated with any other method in an SEM; this represents one of the most difficult samples to characterize due to carbon's low scattering cross section.[11] To generate grain orientation maps of graphene, one could resort to performing conventional DF transmission electron microscopy (TEM).[12,13] Due to DMD technology, however, we can perform this characterization in widely available and accessible SEMs.

2. DETECTOR CONSTRUCTION

A preliminary description of the detector's basic optical design and construction was given previously,[8] and a followup manuscript described the incorporation of the detector into an SEM and showed proof-of-principle data on several material samples.[9] Here we provide a brief overview of the detector with an emphasis on describing the light collection efficiency as it directly influences the overall detector quantum efficiency (DQE) of the p-STEM detector.*

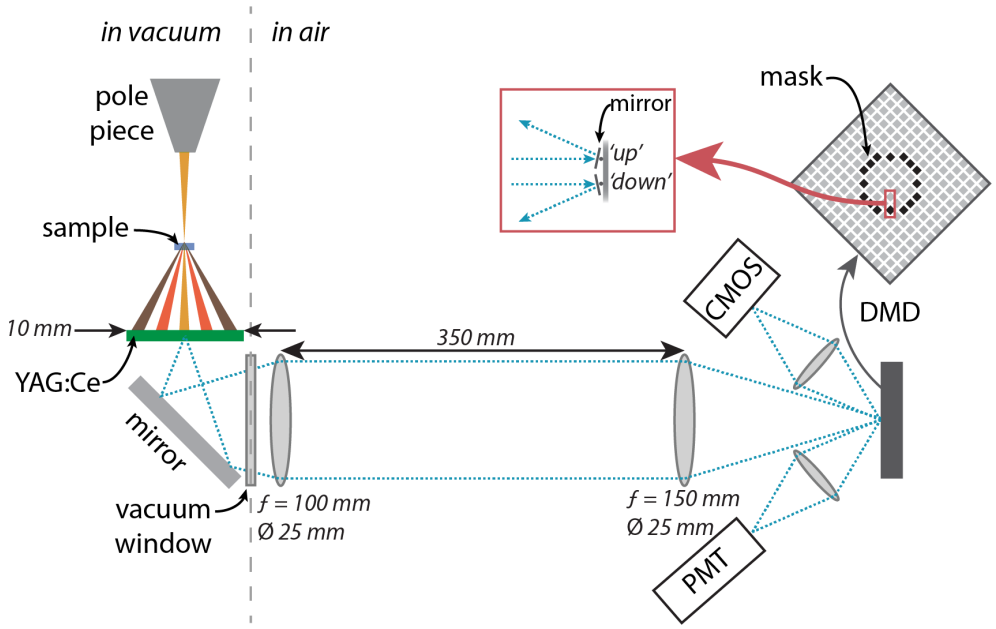


Figure 2. Schematic of the programmable scanning transmission electron microscopy (p-STEM) detector. A focused electron beam is scanned across a thin, electron transparent sample. The transmitted electrons strike a YAG:Ce crystal (Crytur, $t = 100\,\mu m$, ϕ12.7 mm, 50 nm Al top-coat, AR bottom-coat) which emits photons. These photons are imaged out of the vacuum chamber through a window (Eksma Optics, 220-1203M+ARB625) and onto a DMD (Vialux, V-7000) with an achromat pair (Thorlabs, AC254-100-A, AC254-150-A). The upper arm images the DMD surface to a CMOS camera (IDS, UI-3252LE) positioned at the Scheimpflug image plane. The lower arm images (Thorlabs, MAP052550-A) the DMD surface to a PMT (Hamamatsu, H10721-210) which is amplified (Stanford Research, SRS570) before being sent into the auxiliary detector input of an SEM (Zeiss, GeminiSEM 300). The SEM, CMOS camera, and DMD are controlled via software developed in-house.

*Commercial instruments, equipment, or materials are identified only in order to adequately specify certain procedures. In no case does such an identification imply recommendation or endorsement by NIST, nor does it imply that the products identified are the best available for the purpose.

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

2.1 Detector Hardware

The electron detector is constructed as shown in Figure 2 and can be separated into three parts: photon generation, photon transport, and photon detection. Photons are generated when the transmitted electrons strike a YAG:Ce scintillator crystal. Monte Carlo simulations (not shown) can provide insight into the photon generation process.[14] For example, simulations have shown that at a 30 keV beam energy, approximately 15 % of incident electron kinetic energy is lost due to electrons backscattered out of the scintillator. The remaining energy is deposited on the order of 1 μm deep into the scintillator material, far from the aluminum coating which might quench the excited Ce^{3+} ions. The conversion efficiency given for the YAG:Ce is 30 photons per keV ($\lambda = 547$ nm) which gives a yield of 765 photons per electron which we assume are isotropically emitted.

While the photon generation step has a large intrinsic gain, the photon transport from the scintillator to the detector has a significant loss. The dominant loss is due to the small solid angle collected by the imaging optics. Only approximately 1 in 1000 photons is collected, which effectively cancels the gain of the electron generation step. An electron transparent, optically reflective coating (50 nm of aluminum) on the top surface of the YAG:Ce crystal enhances the photon collection efficiency, while the reflective losses in the optical system slightly degrade the photon collection efficiency. In the final step, the photons are detected by a photomultiplier tube, which has a low quantum efficiency (QE) at visible wavelengths. These steps and their relative gain/loss coefficients for the detector configuration used here are given in Table 1.

Table 1. Steps in the photon generation, transport, and detection chain with calculated gain and loss coefficients. The total efficiency is the product of all the coefficients.

step	**coefficient**	**notes**
deposition of energy into the scintillator	0.85	Loss due to electron backscattering estimated from Monte Carlo simulation.[14]
energy to photon conversion	765	Assumes 30 photons ($\lambda = 547$ nm) generated per 1 keV energy deposited with a 30 keV electron beam. Value from manufacturer.[15]
increase in effective source intensity due to reflective coating	1.9	Assumes a 90 % reflectance from 50 nm Al.[16]
fraction of solid angle collected	9.7×10^{-4}	Assumes the ϕ 22.9 mm clear aperture of a $f = 100$ mm lens is the limiting aperture. Includes the change in refractive index going from YAG:Ce ($n = 1.82$) to air ($n = 1$).
reflective losses	0.9	Assumes 99 % transmission (reflectivity) from the various glass (mirror) surfaces. Lenses and vacuum window have antireflective coatings.
loss from DMD	0.65	Includes DMD fill factor, reflectivity, window transmission, and diffraction efficiency as quoted by the DMD manufacturer.[17]
photomultiplier quantum efficiency (QE)	0.12	As quoted from PMT manufacturer.[18]

Combining the gain/loss coefficients, we calculate that for every electron that strikes the scintillator there is a 8.4 % chance that it is detected, which gives a DQE of 0.078 assuming the detector has no noise sources other than counting statistics.[19] Table 1 suggests there are two main routes to improve detector efficiency – the detector QE and the photon efficiency. First, the QE of the detector is 0.12, which leaves significant room for improvement. For example, a GaAsP PMT achieves *ca.* 40 % QE at 550 nm,[20] though for some dark-field imaging applications the DQE could be degraded by the relatively high dark count rate of such detectors. Second, and most important, the collection efficiency of the optics is poor – on the order of 0.1 %. To improve this, larger numerical aperture lenses could be used, albeit at the expense of either increased aberration/distortion and/or reduced field of view.

Experimentally we estimated the probability of electron detection as 0.035 via a photon counting approach. Briefly, at a very low electron beam current the beam current is measured with a Faraday cup. Under effectively identical conditions, the photon count rate was also measured. For low DQEs, the ratio of the photon count rate to electron beam current in electrons per second yields the average number of photons detected per electron incident on the scintillator. The dominant source of error arises from the the difficulty in detecting a photon event in the PMT signal. By adjusting the photon detection criteria within a reasonable range, we estimate an uncertainty of at least to 20 % on the measured probability of electron detection.

The values in Table 1 result in an average photon yield approximately twice that of the photon counting measurement, though this is not unexpected. The calculation assumes the literature/manufacturer values are accurate for the scintillator conversion efficiency and PMT QE, but theses values are known to vary from batch to batch, and the conditions under which they were determined are not specified. Additionally, the calculation assumes the optical system is perfectly aligned, which is likely not the case due to the difficulty of aligning an optical system with the physical constraints imposed by the vacuum chamber. In any case, the calculated and measured detection probabilities are in qualitative agreement, suggesting that the limitations of the optical assembly are understood. We note that we do not characterize the CMOS camera arm of the optical assembly, because the integration times for diffraction patterns are generally many orders of magnitude longer than the dwell times used for SEM image acquisition.

Finally, it is worth placing the DQE measurements/calculations in the context of an actual DF imaging experiment in an SEM. The beam current on a field emission SEM is on the order of 300 pA which corresponds to 2×10^9 electrons per second. For monolayer graphene, a kinematic scattering calculation predicts ≈ 0.2 % of the beam will scatter into the 2nd order diffraction spots. For a DF image based on the 2nd order diffraction spots, the scattering efficiency, and the DQE, result in an electron detection rate of *ca.* 2.9×10^5 electrons per second. This count-rate is notable for two reasons. First, the PMT selected for the p-STEM detector has a dark count rate several orders of magnitude below this expected signal, and thus should not add additional noise. Second, because of the high contrast intrinsic to DF images, a 1 megapixel DF image should be able to be acquired on the order of minutes with a signal to noise in excess of the Rose criterion. In this manuscript, integration times for single DF images were ≤ 120 s, in reasonable agreement with this estimate.

2.2 Detector Software

To utilize the p-STEM detector in an experiment, software was developed to control the SEM, the DMD, and the two optical detectors (PMT and CMOS camera) simultaneously. Two main modes were developed. In 'diffraction mode' the user can load an SEM micrograph into a graphical user interface (GUI) and select several regions of the sample from which to collect diffraction patterns. In this mode, the program tilts all the micromirrors towards the CMOS camera, positions the electron beam at the locations on the sample specified by the user, and then collects and saves the diffraction patterns. These diffraction patterns are then corrected for optical imaging distortions using an affine transform and presented to the user.

In 'imaging mode', the user loads in diffraction patterns from regions on their sample. Then using the GUI, the user can create various digital masks to apply to the DMD which will serve as a digital 'objective aperture' to collect bright and dark field images of the sample. The types of masks that are currently included in the GUI are circular, annular, two dimensional lattices, user selected spots, and combinations thereof. More complex masks can be generated via external scripts. Once the user has specified a diffraction mask, the mask is mapped to the DMD image plane and applied to the DMD. The user can then collect images with the SEM software using the output of the PMT as the signal source. Further details on the method of mapping the diffraction images and masks to different image planes is given in a previous manuscript.[9]

3. APPLICATIONS OF THE DETECTOR TO GRAPHENE

3.1 Orientation Contrast in Monolayer Graphene

As a proof-of-principle we use graphene samples on a lacey carbon support (Ted Pella, 21710) to show orientation-based contrast with the p-STEM detector. These films consist of continuous, mostly flat monolayer graphene across the TEM grid with small patches of bilayers. Folds and wrinkles of the graphene are also present. Coating

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

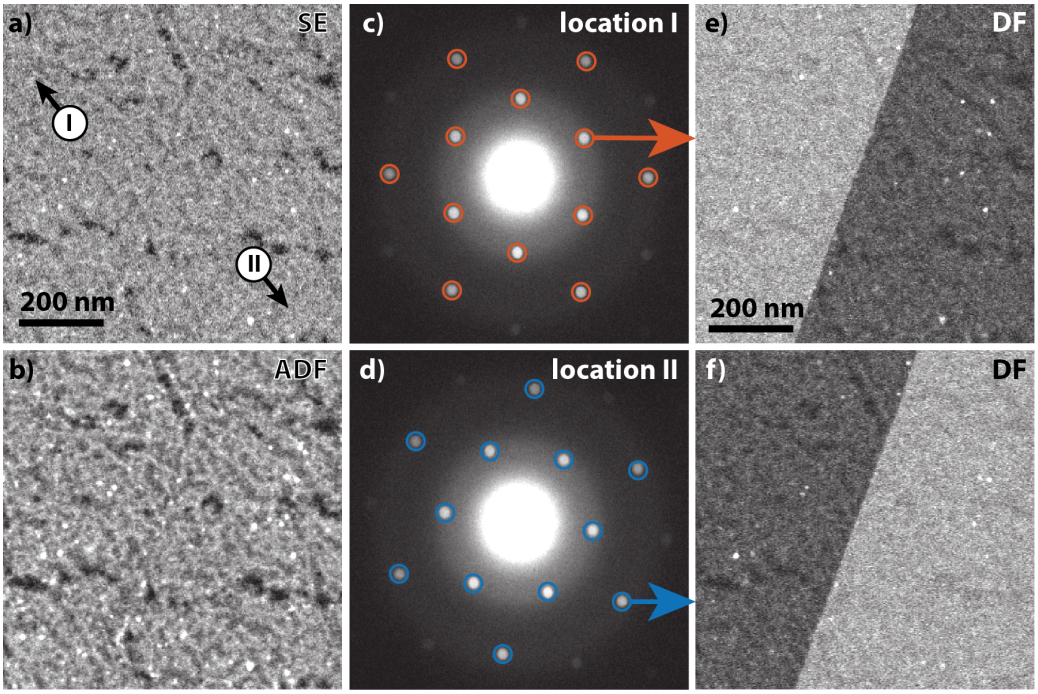


Figure 3. This figure shows how the p-STEM detector can yield orientation contrast from graphene samples that is totally absent from conventional SEM imaging modes. All real-space images are from the same field-of-view. (a) A SE image of a monolayer graphene sample. (b) An ADF image. Neither of these 'conventional' SEM imaging modes show any measurable orientation contrast from the graphene – they only show debris on the graphene surface. (c,d) Diffraction patterns from the two locations specified in (a). The two diffraction patterns are rotated 23° with respect to one another. (e,f) DF images collected by masking the diffraction patterns as shown in (c,d).

the graphene films is a mixture of amorphous (polymer leftover from a transfer step) and crystalline (assumed to be leftover salts from the etching and lift-off process) materials.

Figure 3 shows conventional SEM imaging modes in addition to diffraction contrast images made possible by the p-STEM detector. Figure 3a is an SE image taken with a conventional Everhart-Thornley detector, and Figure 3b is an ADF image taken with an annulus that is set to accept the signal from (transmitted) electrons diffracted by the graphene lattice. Both images show similar features, though the ADF image has greater signal to noise. The small dark patches are likely clean graphene, while the grey texture that coats the surface is assigned as adsorbed organic – likely polymer residue.[21] The small bright spots ($\approx$ 10 nm) are crystalline debris.

Figure 3c-d show diffraction patterns collected in 'diffraction mode' at locations I and II as specified in Figure 3a. The diffraction patterns show two main features. First, they show the first three diffraction orders from graphene as spots (the 3rd order is extremely weak as expected). The relative intensity of the 1st and 2nd order diffraction spots was measured to be close to 1 and indicates that the graphene is monolayer thickness.[22] The two diffraction patterns show a 23° relative orientation. Second, low intensity diffuse rings are visible near the 1st and 2nd order diffraction spots. These rings are due to the amorphous debris coating the sample. Prolonged exposure to the electron beam will increase these diffuse rings via electron-beam-induced sample contamination.

Figure 3e-f show DF images collected by creating two different DF masks that collect the diffraction spots of each diffraction pattern in Figure 3c-d. In these images, there is a clear contrast between the left and right sides of the field of view separated by a distinct interface. This interface is a grain boundary, and the spatial resolution is limited only by that of the SEM image. We note that this interface is not visible in conventional

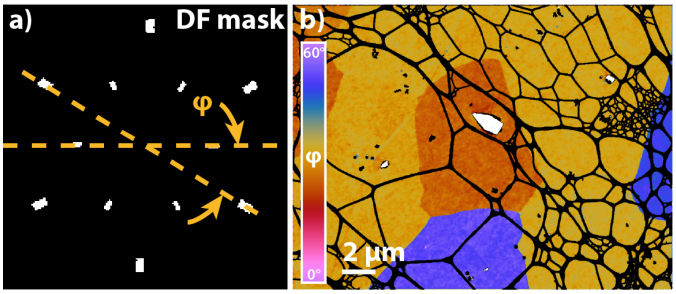


Figure 4. This figure shows how the p-STEM detector can perform orientation mapping of graphene. (a) A DF mask is designed to overlap with the 1st and 2nd order graphene diffraction spots over a finite angular range of the in-plane orientation angle, ϕ. A series of 30 such masks are generated for $\phi \in \{0°, 2°, 4°...58°\}$ and a DF image is collected for each mask. From this stack of images, an orientation map can be constructed and is displayed in (b). Here the lacey carbon is colored black, and holes in the graphene sheet are colored white. Several grains with different orientation are obvious in the field of view.

SEM imaging modes (BF, ADF, SE) regardless of the image processing steps taken. The data show that the p-STEM detector is capable of obtaining image contrast not available with conventional SEM imaging modes.

Once this diffraction contrast is observed, it is relatively straightforward to develop an automated method for mapping the orientation of graphene domains. Here, a series of digital masks is generated that has the symmetry of the graphene diffraction pattern, but with each mask having a different in-plane orientation parameterized by ϕ (Figure 4a). By generating a series of these masks that span the angular range $\phi \in [0°, 60°]$ and collecting a DF image for each mask, we can determine the orientation of the graphene lattice at each pixel. Figure 4b shows an orientation map with several grains of graphene visible. To obtain high quality orientation maps, we performed rigid image registration to account for sample drift.[23] Additionally, the data in Figure 4b was filtered with a total variational denoising algorithm on each DF image prior to computing the orientation. Total variation denosing significantly reduces the noise in piecewise constant images without smoothing the edges (i.e. grain boundaries).[24] In the present work, we used a first circular moment analysis of the image stack to obtain the graphene orientation.[25] More advanced peak finding or pattern matching methods[26] would likely be more accurate at an increased computational cost.

3.2 Orientation Contrast in Bilayer Graphene

In addition to mapping the orientation contrast in monolayer graphene films, this detector can be used to map the orientation in bilayer films. Figure 5a shows a SE image of a graphene film. Diffraction patterns were collected at location I and II as indicated and are shown in Figure 5b-c respectively. Location II shows only a single orientation of graphene, while location I shows a two orientations of graphene. Three DF images were collected corresponding to the masks shown. The DF image corresponding to the dotted blue mask in Figure 5b represents a background image due to electrons scattered from the surface debris and lacey carbon. Figure 5e-f correspond to the two DF masks that each collected the electrons scattered from a particular graphene lattice orientation. The data show that there is an annular shaped island of graphene on top of a continuous graphene sheet.

3.3 Moire Fringe Contrast in Multilayer Graphene

Multiple layers of graphene with nearly the same orientation will give rise to moire fringes/patterns in DF images. If visible, these periodic signals give insight into the local stacking of graphene multilayers. Figure 6a shows a diffraction pattern collected from a multilayer region of graphene. Three DF images were collected (Figure 6b-d) using the three diffraction spots specified. In all three DF images, moire fringes are readily visible signifying strain and/or misorientation between the layers.[27]

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

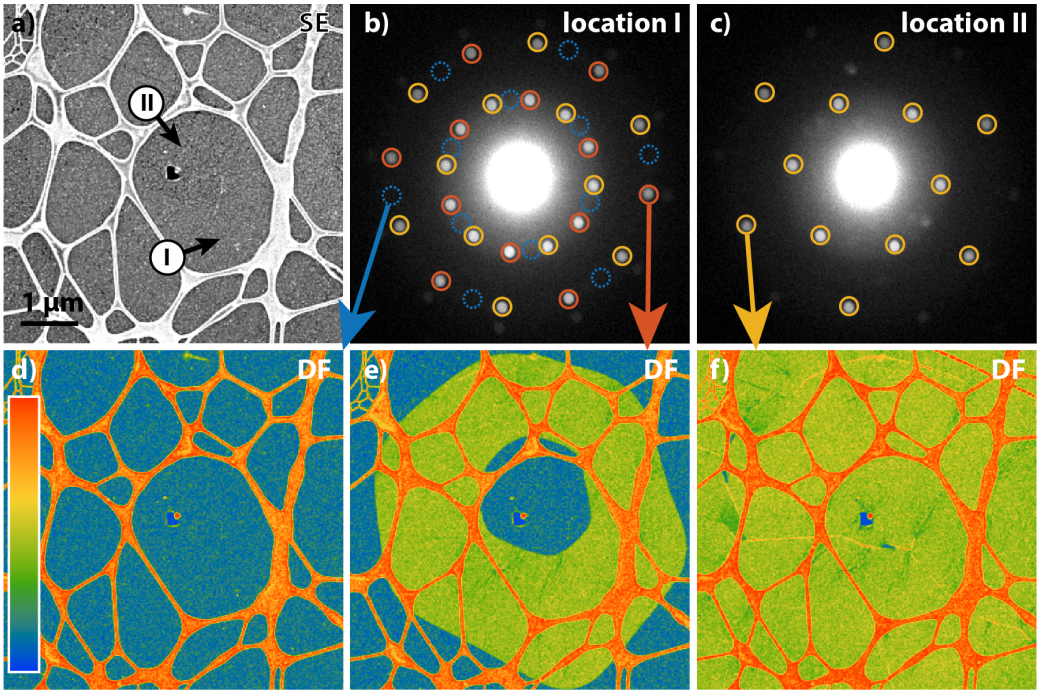


Figure 5. This figure shows how the p-STEM detector can characterize the orientation of graphene bilayers. (a) A SE image of a region of a graphene film supported on lacey carbon. (b,c) Diffraction patterns collected at the two locations specified in (a). Location I shows two graphene diffraction patterns with a relative orientation of 30°, while location II only shows one of the two graphene orientations. (d,e,f) DF images collected by masking the diffraction patterns as shown in (b,c). Since the DF mask used to collect (d) is not coincident with a graphene lattice, it serves as a background image due to scattering from surface debris and the lacey carbon. Here the lacey carbon is colored red, the surface debris is colored blue. In (e,f) graphene shows up as green due to the fact the the mask is capturing the electrons scattered by the graphene lattice. These images show that while the conventional SE image is not sensitive to the layer number and orientation of the graphene lattice, the DF images can easily discern the presence of a rotated bilayer island on a contiguous graphene monolayer support.

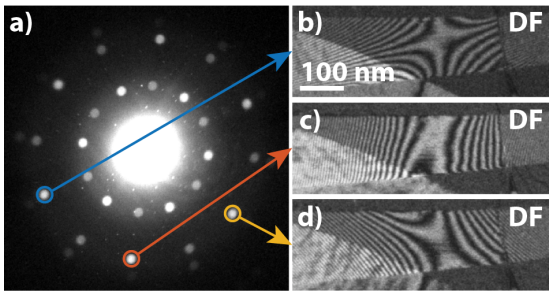


Figure 6. This figure highlights that the DF imaging afforded by the p-STEM is sensitive to lattice strain. (a) A diffraction pattern is collected from the field of view in (b,c,d). (b,c,d) DF images collected by masking a single 2nd order graphene diffraction spot as specified in (a). Moire fringes are readily visible and contain information on the local stacking of graphene multilayers.

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

4. CONCLUSIONS

Traditional imaging modes in an SEM have limited sensitivity to the crystallographic orientation of 2D materials. Herein, crystallographic contrast was not observed on monolayer graphene when using an Everhart Thornley SE detector or a BF or ADF transmission electron detector. The basic problem is that orientational contrast requires a non-axially symmetric detector capable of differentiating in-plane rotations. One method to construct a detector sensitive to in-plane orientation uses a DMD in an image plane conjugate to the diffraction pattern to serve as a programmable DF mask. By incorporating DMD technology into the p-STEM detector, we are able to switch between 'imaging' and 'diffraction' modes without any macroscopic moving parts. This allows the user to perform all DF masking operations digitally. We demonstrate that the p-STEM detector has the potential to characterize grain orientation in 2D materials. We believe that DMD technology will enable the SEM to be used to investigate phenomena that previously required the use of a TEM.

ACKNOWLEDGMENTS

We thank Ryan White (NIST) and Ann Chiaramonti Debay (NIST) for helpful microscopy related discussions, and Callie Higgins (NIST) for discussions related to optics. This work was performed while B.W.C. held a National Research Council Postdoctoral Associateship. The authors also thank the NIST Small Business Innovation Research Program for supporting RadiaBeam Technologies' detector development under Contract SB1341-14-CN-0026.

REFERENCES

[1] Goldstein, J. I., Newbury, D. E., Echlin, P., Joy, D. C., Lyman, C. E., Lifshin, E., Sawyer, L., and Michael, J. R., [*Scanning Electron Microscopy and X-ray Microanalysis*], Springer US, Boston, MA (2003).

[2] Everhart, T. E. and Thornley, R. F. M., "Wide-band detector for micro-microampere low-energy electron currents," *Journal of Scientific Instruments* **37**, 246–248 (jul 1960).

[3] Zhou, Y., Fox, D. S., Maguire, P., O'Connell, R., Masters, R., Rodenburg, C., Wu, H., Dapor, M., Chen, Y., and Zhang, H., "Quantitative secondary electron imaging for work function extraction at atomic level and layer identification of graphene," *Scientific Reports* **6**, 21045 (aug 2016).

[4] Yoo, Y., Degregorio, Z. P., and Johns, J. E., "Seed Crystal Homogeneity Controls Lateral and Vertical Heteroepitaxy of Monolayer MoS2and WS2," *Journal of the American Chemical Society* **137**(45), 14281–14287 (2015).

[5] Holm, J. and Keller, R. R., "Angularly-selective transmission imaging in a scanning electron microscope," *Ultramicroscopy* **167**, 43–56 (aug 2016).

[6] Faruqi, A. and McMullan, G., "Direct imaging detectors for electron microscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **878**, 180–190 (jan 2018).

[7] Fundenberger, J. J., Bouzy, E., Goran, D., Guyon, J., Yuan, H., and Morawiec, A., "Orientation mapping by transmission-SEM with an on-axis detector," *Ultramicroscopy* **161**, 17–22 (2016).

[8] Jacobson, B. T., Gavryushkin, D., Harrison, M., and Woods, K., "Angularly sensitive detector for transmission Kikuchi diffraction in a scanning electron microscope," *Proceedings of SPIE of SPIE* **9376**(310), 93760K (2015).

[9] Caplins, B. W., Holm, J. D., and Keller, R. R., "Transmission imaging with a programmable detector in a scanning electron microscope," *Ultramicroscopy* **196**, 40–48 (jan 2019).

[10] Cowley, J. and Spence, J., "Innovative imaging and microdiffraction in stem," *Ultramicroscopy* **3**, 433–438 (jan 1978).

[11] Reimer, L., [*Scanning Electron Microscopy*], vol. 45 of *Springer Series in Optical Sciences*, Springer Berlin Heidelberg, Berlin, Heidelberg (1998).

[12] Kim, K., Lee, Z., Regan, W., Kisielowski, C., Crommie, M. F., and Zettl, A., "Grain Boundary Mapping in Polycrystalline Graphene," *ACS Nano* **5**, 2142–2146 (mar 2011).

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

[13] Huang, P. Y., Ruiz-Vargas, C. S., van der Zande, A. M., Whitney, W. S., Levendorf, M. P., Kevek, J. W., Garg, S., Alden, J. S., Hustedt, C. J., Zhu, Y., Park, J., McEuen, P. L., and Muller, D. A., "Grains and grain boundaries in single-layer graphene atomic patchwork quilts," *Nature* **469**, 389–392 (jan 2011).

[14] Demers, H., Poirier-Demers, N., Couture, A. R., Joly, D., Guilmain, M., De Jonge, N., and Drouin, D., "Three-dimensional electron microscopy simulation with the CASINO Monte Carlo software," *Scanning* **33**(3), 135–146 (2011).

[15] CRYTUR, "YAG:Ce Technical Parameters."

[16] Hass, G. and Waylonis, J. E., "Optical Constants and Reflectance and Transmittance of Evaporated Aluminum in the Visible and Ultraviolet," *Journal of the Optical Society of America* **51**, 719 (jul 1961).

[17] Instruments, T., "DLP7000 Technical Documents."

[18] Hamamatsu, "Photosensor module: H10721-210."

[19] Browne, M. and Ward, J., "Detectors for stem, and the measurement of their detective quantum efficiency," *Ultramicroscopy* **7**, 249–262 (jan 1982).

[20] Thorlabs, "PMT2101 - GaAsP Amplified PMT."

[21] Dyck, O., Kim, S., Kalinin, S. V., and Jesse, S., "Mitigating e-beam-induced hydrocarbon deposition on graphene for atomic-scale scanning transmission electron microscopy studies," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* **36**, 011801 (jan 2018).

[22] Shevitski, B., Mecklenburg, M., Hubbard, W. A., White, E. R., Dawson, B., Lodge, M. S., Ishigami, M., and Regan, B. C., "Dark-field transmission electron microscopy and the Debye-Waller factor of graphene," *Physical Review B* **87**, 045417 (jan 2013).

[23] Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R., "Efficient subpixel image registration algorithms," *Optics Letters* **33**, 156 (jan 2008).

[24] Rudin, L. I., Osher, S., and Fatemi, E., "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena* **60**, 259–268 (nov 1992).

[25] Mardia, K. V. and Jupp, P. E., eds., [*Directional Statistics*], Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, USA (jan 1999).

[26] Chen, Y. H., Park, S. U., Wei, D., Newstadt, G., Jackson, M. A., Simmons, J. P., De Graef, M., and Hero, A. O., "A Dictionary Approach to Electron Backscatter Diffraction Indexing," *Microscopy and Microanalysis* **21**, 739–752 (jun 2015).

[27] Brown, L., Hovden, R., Huang, P., Wojcik, M., Muller, D. A., and Park, J., "Twinning and Twisting of Tri- and Bilayer Graphene," *Nano Letters* **12**, 1609–1615 (mar 2012).

Caplins, Benjamin; Holm, Jason; Keller, Robert.
"A programmable dark-field detector for imaging two-dimensional materials in the scanning electron microscope."
Paper presented at SPIE Photonics West 2019, San Francisco, CA, United States. February 2, 2019 - February 7, 2019.

A Phish Scale: Rating Human Phishing Message Detection Difficulty

Michelle P. Steves
National Institute of Standards and Technology
michelle.steves@nist.gov

Kristen K. Greene
National Institute of Standards and Technology
kristen.greene@nist.gov

Mary F. Theofanos
National Institute of Standards and Technology
mary.theofanos@nist.gov

***Abstract*—As organizations continue to invest in phishing awareness training programs, many Chief Information Security Officers (CISOs) are concerned when their training exercise click rates are high or variable, as they must justify training budgets to those who question the efficacy of training when click rates are not declining. We argue that click rates should be expected to vary based on the difficulty of the phishing email for a target audience. Past research has shown that when the premise of a phishing email aligns with a user's work context, it is much more challenging for users to detect a phish. Given this, we propose a Phish Scale, so CISOs and phishing training implementers can easily rate the difficulty of their phishing exercises and help explain associated click rates. We based our scale on past research in phishing cues and user context, and applied it to previously published data and new data from organization-wide phishing exercises targeting approximately 5000 employees. The Phish Scale performed well with the current phishing dataset, but future work is needed to validate it with a larger variety of phishing emails. The Phish Scale shows great promise as a tool to help frame data sharing on phishing exercise click rates across sectors.**

Keywords—phishing cues, embedded phishing awareness training, operational data, network security, phishing defenses, security defenses

I. Introduction

According to Cybersecurity Ventures' 2017 Official Annual Cybercrime Report, it is estimated that cybercrime damages will cost the world $6 trillion annually by 2021 [5]. These cost projections are supported by historical cybercrime figures and recent year-over-year growth. Furthermore, there has been a notable increase in hacking activities sponsored by hostile nation states, as well as activities from organized crime syndicates. Finally, the cyber attack surface continues to grow, in large part due to an explosion of Internet of Things (IoT) devices. Humans are another particularly important component of the overall attack surface, as social engineering continues to be successful. In recognition of the importance of human behavior in cybersecurity, organizations are more widely investing in cybersecurity awareness programs for their computer users, and often on phishing training in particular. Embedded phishing awareness training is popular—and in some cases, mandated—in a wide variety of sectors, such as financial services, government, healthcare, and academia. In this type of training, simulated phishing emails are sent that mimic real-world threats, in order to raise employee phishing awareness.

Workshop on Usable Security (USEC) 2019
24 February 2019, San Diego, CA, USA
ISBN 1-891562-57-6
https://dx.doi.org/10.14722/usec.2019.23028
www.ndss-symposium.org

Not surprisingly, many Chief Information Security Officers (CISOs) are concerned when their training exercise click rates are high. This is especially true for more mature or long-running awareness programs, as CISOs often expect lower and lower click rates to show the effectiveness of training. Further, the Return on Investment (ROI) for such training may be questioned if click rates are high or even variable. However, low click rates do not necessarily indicate training effectiveness and may instead mean the phishing emails used were: 1) too easy, 2) not contextually relevant for most staff, or 3) the phish was repeated or very similar to previous exercises. In fact, low click rates and training programs in general, can generate a false sense of security or complacency if considered in isolation. Phishing awareness training program click rates must be part of a more comprehensive, metrics-informed approach to effectively understand and combat phishing threats [15].

Past work [14] has shown that click rates will vary based on the contextual relevance of the phish, with highly contextually relevant phish resulting in extreme spikes in click rates—despite years of phishing awareness training. Furthermore, attackers continue to refine and vary phishing attack premises. Although "traditional" phishing emails are still quite successful, attackers are becoming more sophisticated and creative all the time. Additionally, there is a treasure trove of readily available information online that attackers can use to better tailor phish and capitalize on contextual relevance. While some information is willingly and openly shared by users on social media, much other information has been exposed through large-scale data breaches, such as the recent Facebook hack [29].

While repetition is important for training phishing recognition and for conditioning reporting behavior, simple repetition of the same or very similar phishing emails does not represent the full spectrum of phishing threats observed in the real world. It is important to vary phishing exercises appropriately and challenge staff with contextually relevant phish of varying difficulty to provide training on new scams—for which variable click rates should be expected. This should not be viewed as a negative effect but rather a positive outcome, as it means organizations are truly training their staff with phish that represent current real-world threats. But how exactly does one measure the difficulty of a given phishing email? While we can certainly measure click rates post-hoc and infer detection difficulty somewhat from those numbers, we would prefer an a priori method of difficulty determination. In discussions with CISOs at [17], [18], and others, we found that a method to determine phishing message difficulty would indeed be highly beneficial for those responsible for phishing training implementation. To meet this need, we propose a Phish Scale,

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

an easy way for CISOs and training implementers to characterize the difficulty of their phishing exercises and provide context for the associated metrics. This context is a missing element that training implementers need to improve the training benefit of their exercises and subsequent ROI.

In this paper, we describe our exploratory effort to construct a preliminary conceptual version of the Phish Scale and its components. Further, we use the Phish Scale to determine the difficulty rating for each of seven real-world phishing training exercises that are described in detail. Then we observe if the Phish Scale difficulty rating for each exercise aligns with the exercise's actual click rate. Finally, we discuss our observations, limitations of the effort to-date, as well as future work including refinement of its components and validation of the overall scale through a wider variety of phishing emails.

II. Background

Technological and human-centered approaches are used in conjunction to combat email phishing. Technologically-focused approaches include mechanisms like filtering, firewalls, and blacklists, whereas human-centered approaches tend to focus on cybersecurity awareness training, and often on phishing specifically. Due in part to advances in Machine Learning (ML) and Artificial Intelligence (AI), email filters in particular, are becoming ever more effective at blocking generic spam. This has meant that users now see fewer emails of this nature in their inboxes. Recent work has posited the existence of the "Prevalence Paradox" [34], suggesting that users may therefore be more vulnerable when such emails do get through, due to their reduced experience with potentially malicious emails. Yet other work [25] has shown that people often expand their concept of a given stimulus in response to a decrease in the prevalence of said stimulus, for example, seeing neutral faces as threatening when threatening faces became rare. Although the series of experiments by Levari et al. did not address phishing specifically, given the set of topics they investigated, it would certainly be plausible to expect their findings to hold in the phishing domain. We hope additional research on the effects of prevalence on phishing detection—for both humans and AI—will reconcile different findings on prevalence.

In addition to prevalence, there are numerous other factors that complicate human detection of phishing emails. There are several existing theories and models of phishing susceptibility that are highly relevant for the development of a Phish Scale. These theories and models directly address the types of email cues, tactics, and individual user characteristics that together help—at least partially—explain the relative ease or difficulty of human phishing detection.

Protection Motivation Theory, or PMT [33] addressed user perceptions of threat and corresponding perceived threat management ability. PMT has largely been applied to security behavior in general, although Wang et al. [41] did apply PMT specifically to phishing threat perception. Much more recently than PMT, which was originally proposed in 1975, an Integrated Information Processing Model of Phishing Susceptibility, or IIPM, was proposed [38]. The IIPM proposed that users' limited attentional resources for information processing are essentially hijacked when certain techniques like urgency are used to influence behavior, meaning that users rely on heuristic information processing (System I, [22]), rather than engaging in deeper, more systematic processing (System II, [22]). When this type of surface level information processing style is used it makes users more likely to overlook or ignore cues that might otherwise tip a user off as to the legitimacy of the email, such as an incorrect sender address. In 2016, Vishwanath et al. proposed the Suspicion, Cognition, and Automaticity Model, or SCAM, which posited that individual user characteristics cause variability in the use of heuristic processes for email evaluation [39].

Recent work by Williams, Hinds, and Joinson [42] considered these three models (PMT, IIPM, and SCAM) within the work context of an international organization with sites in the UK, finding that the presence of authority cues increased the likelihood that users would click a suspicious email link [42]. In addition to the types of models or theories such as PMT, IIPM, SCAM, there is a large wealth of prior work investigating or describing the impact of particular email cues, such as inclusion of authority and urgency cues. Research on phishing cues is particularly relevant for development of a Phish Scale, as email users rely on cues to determine if a particular email message is a phish.

Indeed, anti-phishing advice and training stress the characteristics of phishing messages that email users should look for; these are often called cues, indicators and hooks. The list of cues is long and varied, such as those contained in [26], [30]. Because there is no set pattern of which cues may be contained in any particular message, the task for users when determining if a message is a phish is harder than if the list were very short. Making the task even more difficult, prior work shows the alignment of the phish's premise and user context affects which cues the user finds to be salient. Further, the same cue can be compelling for some users but suspicion generating for others—depending on the user's context [14], [15].

In the Greene et al. [14] study, phishing exercise data were collected over 4.5 years in an ecologically valid workplace setting, with corresponding survey data for the final year. The study found that user context was extremely important in phishing susceptibility; the authors proposed that it was the lens through which users viewed and interpreted email cues. When a user's work context was misaligned with the premise of the phishing email, they were more likely to attend to suspicious cues. For example, they have no invoicing responsibilities at work and the phishing email was purportedly an unpaid invoice. In contrast, when a user's work context was well aligned with the phishing email premise, they were more likely to attend to compelling cues, and completely ignore or largely discount suspicious cues. In this case, if the user is directly responsible for paying invoices at work and the phishing email was purportedly an unpaid invoice.

Greene et al. [14] emphasized the importance of phishing research in the workplace setting, as much prior phishing work was conducted in laboratories with artifical user contexts or university settings that can be quite different than the workplace. Williams et al. [42] also addressed this need for workplace data in their research. One of the few other studies situated in the workplace was conducted by Caputo et al. [4], but due to limitations was only able to suggest the possible importance of user context. We further contribute to the growing corpus of workplace-based phishing research, by applying our Phish Scale to three previously published workplace-based phishing

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

exercises in [14], as well as, reporting on and applying the Phish Scale to four new workplace-situated phishing exercises. Two of our new exercises have much larger sample sizes, with *n*'s of ~5000 for each exercise, compared to those previously reported in [14], with *n*'s of ~70 for each exercise.

III. METHOD

To assist those tasked with implementing phishing awareness training programs, it is important to consider the relative detection difficulty of training messages. Phishing messages, whether those intended for training or actual threats, can be more or less difficult *for a given work group* to detect as a phishing attempt. Understanding the detection difficulty helps phishing awareness training implementors in two primary ways: 1) by providing context regarding training message click and reporting rates for a target audience, and 2) by providing a way to characterize actual phishing threats so the training implementor can reduce the organization's security risk by tailoring training to the types of threats their organization is facing. To this end, we are attempting to develop a Phish Scale—to help practitioners rate training messages both to contextualize click rates for embedded phishing awareness training as well as tailor training efforts. We anticipate it will provide CISOs with another metric to help gauge the progress of their awareness programs over time and address risk. The scale is intended to categorize the detection difficulty of a phishing message with respect to a target audience.

In the remainder of this section, we describe the Phish Scale and the operationalization of its components into a single framework. In the next section, we present data from seven workplace-situated phishing awareness training exercises to illustrate how to derive a phish difficulty rating using the Phish Scale. The data were gathered with appropriate human subjects approval at NIST.

A. The Phish Scale

To develop our Phish Scale, we began by considering the primary elements that CISOs and/or training implementors use when selecting and customizing phishing training exercises. These elements are scenario premise and message content. The scenario premise may pertain to a relatively new threat or an older threat that remains effective for a particular target audience. The message content is often customizable by the trainer and contains the cues that trainees might use to detect the training phish. For this exploratory effort, we root the Phish Scale in these two primary elements: the *cues contained in the message* and the *premise alignment for the target audience*.

Other factors such as personality, phishing tactics knowledge, concern for security, concern for consequences, and the like certainly affect click rates, and ultimately we intend to consider incorporating additional factors such as these; we return to this topic in the future work section. However, for now, this effort starts with message cues and premise alignment as these elements undoubtedly play crucial roles in phishing detection by humans and, importantly, they can be categorized by training implementors for a given target audience. For this initial effort at characterizing detection difficulty, the Phish Scale components are:

1) A rating system for observable characteristics of the phishing email itself, such as the number of cues, nature of the cues, repetition of cues, and so on.

2) A rating system for alignment of the phishing email premise with respect to a target audience.

Table I presents our exploratory, conceptual framework illustrating how detection difficulty rating is arrived at once the categories for number of cues and premise alignment are determined. In an attempt to keep the categorization relatively simple for training implementors, we used three categories for each component and assigned labels representing relative ranges for each. Next we discuss the operationalization of each component. In the following section, we walk through marrying the real-world phishing training exercise data with these conceptual categorizations and discuss our observations.

TABLE I: THE PHISH SCALE

Number of Cues	Premise Alignment	Detection Difficulty
Few (more difficult)	High	Very difficult
	Medium	Very difficult
	Low	Moderately difficult
Some	High	Very difficult
	Medium	Moderately difficult
	Low	Moderately to Least difficult
Many (less difficult)	High	Moderately difficult
	Medium	Moderately difficult
	Low	Least difficult

In the conceptual framework we acknowledge the stronger influence of premise alignment component over cues, as reported in [14]. This is reflected in the detection difficulty rating tending to be at the *Very difficult* or *Moderately difficulty* rating when the premise alignment is categorized as *High* or *Medium*. Additionally, there are more *Very difficult* detection difficulty rating assignments than *Least difficult* rating assignments in the entire conceptual Phish Scale framework. The detection difficulty rating for the combination of *Some* cues and *Low* premise alignment was given a range from *Moderately to Least difficult* rating, further reflecting our belief that even a *Low* premise alignment can have a disproportionate effect on increasing detection difficulty. While we expect all of the ratings to be informed with empricial data, this is especially true for this particular combination (low premise alignment and some cues). Finally, we purposefully did not label a category as *Easy to detect* or similar, as we expect that the premise of any phishing message will typically align for *at least a few users* and for them, detection is often not easy.

Ultimately, we anticipate that each detection difficulty rating will equate to a range of click rates. For example, the phishing training messages that have a corresponding detection difficulty rating of *least difficult* may be expected to have a click rate of less than 10 %. We return to this topic in the discussion after we have examined the empirical data presented in the next section.

B. Phishing message cues

To incorporate the effect of phishing message cues in the scale, we decided to use the count of instances of those characteristics that are present in the message being rated. Our reasoning is that the fewer phishing cues present in a message, the more difficult it is to detect. Conversely, the more cues

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

present, the more opportunities for a user to notice a tip-off that generates suspicion. We realize the effect of any single cue or hook can differ from instance to instance and person to person. Indeed, we return to this topic in the discussion. Currently, there are three categories in the framework to describe the quantity of these characteristics: *Few* (fewer opportunities to detect), *Some,* and *Many* (more opportunities to detect).

Before we can count cues, we needed to determine which phishing characteristics—the list of cues, indicators and hooks—are appropriate for inclusion in the framework. From [14] we see that a particular phishing characteristic may either be suspicion-generating (a tip-off) or compelling (a hook), depending on the user's context. In keeping with prior literature, we use the term "cue." However, we mean it in the broader sense of a phishing message characteristic. We require that each cue included in the framework be able to be tied to an objectively observable characteristic in a message.

From the literature we considered the compendiums of phishing cues in [26] and [30]. We used the cues given in [26] as a starting point. Additionally, we modified the categories in an attempt to order the cues from those that are often suspicion-generating, such as errors, to those that are typically compelling, such as common tactics, these tactics being commonly used because they continue to be compelling. This is a rough ordering of categories at best, but we felt it is better suited to counting cues than those given in [26] and [30]. The categories are: *Error*—relating to spelling and grammar errors and inconsistencies contained in the message; *Technical indicator*—pertaining to email addresses, hyperlinks and attachments; *Visual presentation indicator*—relating to branding, logos, design and formatting; *Language and content*—such as a generic greeting and lack of signer details, use of time pressure and threatening language; and, *Common tactic*—use of humanitarian appeals, too good to be true offers, time-limited offers, poses as a friend, colleague, or authority figure, and so on. We wove in additional phishing characteristics from [14], [30], and others.

Table II provides the list of cues we identified that are objectively present in phishing messages. The table in the Appendix A contains the same list of cues, but is expanded with a brief description of each, associated references, and the criteria we used when deciding if a particular cue was observably present in an individual message. To determine the cues count, use the criteria given in Appendix A for each cue, count how many instances for each and sum for a total.

For this initial effort, we recognize this list is not exhaustive and will be expanded. Additionally, we anticipate some form of weighting will be useful to reflect cue saliency. Given the variability in cue saliency for individuals within a target population, this is a non-trivial exercise. These are refinements we expect will come with additional development of the scale.

For the purpose of the Phish Scale, we did not include phishing message cues related to mismatches with the user's world, such as an individual's particular work responsibilities or an individual's expectations, for example expecting an important phone call. Work responsibilities and general workplace expectations for the target audience are folded into the premise alignment component of the Phish Scale.

TABLE II: PHISHING MESSAGE CUES

Cue Type	Cue Name
Error	Spelling and grammar irregularities
	Inconsistency
Technical indicator	Attachment type
	Sender display name and email address
	URL hyperlinking
	Domain spoofing
Visual presentation indicator	No/minimal branding and logos
	Logo imitation or out-of-date branding/logos
	Unprofessional looking design or formatting
	Security indicators and icons
Language and content	Legal language/copyright info/disclaimers
	Distracting detail
	Requests for sensitive information
	Sense of urgency
	Threatening language
	Generic greeting
	Lack of signer details
Common tactic	Humanitarian appeals
	Too good to be true offers
	You're special
	Limited time offer
	Mimics a work or business process
	Poses as friend, colleague, supervisor, authority figure

C. Phishing premise alignment

Incorporating premise alignment is a process of characterizing the pertinence of the email message premise for the target audience. It attempts to capture alignment with the following: work responsibilities and business practice plausibility for the target audience. For example, the organization's current business practices and staff expectations reflecting the organization's workplace culture. The premise alignment is expected to be determined by the training implementor—someone with knowledge of the target audience's work responsibilities and expectations as a group.

We use three categories to characterize the alignment: *High*, *Medium*, and *Low*. To determine premise alignment, the training implementer must understand and categorize the premise relevancy for portions of the target audience using the guidelines below.

1) High alignment

For high premise alignment, there should be a significant portion of the target audience for which the premise matches with work responsibilities, is highly plausible, and/or aligns strongly with an audience-relevant event. For example, if the recipient population is the finance department and the phishing

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

message has a premise of a late/missed payment, the overall alignment is high.

2) Medium alignment

Medium alignment is achieved with either case: a) when the premise has plausible but weak context alignment with a large portion of the target audience or b) when the premise has moderate context alignment with a small portion of the target audience. For example, if the recipient population mostly works in one physical location and the phishing message has a moderately pertinent premise for the few members of the recipient population who work in another physical location.

3) Low alignment

There is low alignment when the premise pertains to a topic that is not relevant or plausible to the target audience. For example, if the recipient population is the finance department and the phishing message premise pertains to a Call for Papers on biotech research or a similarly unrelated topic, the overall alignment is low.

IV. Application of the Phish Scale

In this section we present data from seven phishing training exercises and use the Phish Scale to determine the detection difficulty rating for each exercise. We followed the appropriate human subjects approval process for our institution. All the awareness training exercises were situated in a workplace environment. First, we provide a description of each exercise, its premise alignment rationale, and a brief description of the target audience size and available information with respect to the premise alignment. Then we gather the data in Table III, including the cue counts provided in Appendix B, and show the detection difficulty rating for each exercise side-by-side with the actual click rate.

The first three phishing exercises (new voicemail, unpaid invoice, and order confirmation) were initially reported in Greene et al. [14]. Here we expand their descriptions to count the cues and categorize the premise alignment. The subsequent four phishing exercises (Gmail, weblogs, Valentine, and security token) represent new data.

Note that although we assigned premise alignment category ratings to these exercises—rather than the training implementers—we did so with pertinent input from the training implementers.

1) Phishing exercise descriptions

a) New voicemail

Message description: The new voicemail phish appeared to be from a fictitious CorpVM (corpvm@webaccess-alert.com). It appeared to be a system-generated email, with the subject line reading, "You have a new voicemail." There was a large black and green banner at the top of the email with the text, "CorpVM" in white. There were no logos present in the email, however, there was a small black footer with "© 2015 CorpVM Inc." in white. The body of the email began, "You have a new voicemail!" centered in bold text, followed by, "From: Unknown Caller, Received: 03/06/2016, Length: 00:52." Below that text was a personalized [Firstname Lastname] line, followed by, "You are receiving this message because we were unable to deliver it voice message did not go through because the voicemail was unavailable at that moment. To listen to this message, please click here. You must have speakers enabled to listen to the message. * The reference number for this message is qvfl_cjl09-9107319601-2125579909-62. The length of transmission was approximately 52 seconds. The receiving machine's ID: YJH35-TW410-F37JZL. Thank you." Finally, the email closed with smaller text in italic that read, "This is a system-generated message from a send-only address. Please do not reply to this email."

Premise alignment: The alignment is categorized as Medium—the premise was plausible; around the same time as the exercise, a new business process for voicemail notification, *not* delivery, was being rolled out, although without much fanfare. Even though the premise was plausible, it had no or weak context alignment for most, although not all, of the target audience based on survey feedback reported in [14].

Target audience: One Operational Unit (OU) within NIST, handling financial matters (ordering and invoice reconciliation), administrative program support, and technical program support. $n = 69$

b) Unpaid invoice

Message description: The unpaid invoice phish appeared to be from a fictitious employee of the same institution as the email recipients, a fellow Federal employee named Jill Preston (jill.preston@nist.gov). The subject line was, "Unpaid invoice #4806." The greeting was personalized with "Dear [Firstname Lastname]." The email body said, "Please see the attached invoice (.doc) and remit payment according to the terms listed at the bottom of the invoice. Let us know if you have any questions. We greatly appreciate your prompt attention to this matter!" The email simply closed with the name "Jill Preston." There was no other contact information included below the name. Of note, there was a file extension mismatch between the way the attachment was referred to in the body of the email (as a *.doc*) and the way the attachment itself was labeled, it appeared to be a *.zip*, with the filename, "invoice_S-37644806.zip". The unpaid invoice phish mimicked the Locky ransomware [32], a real-world threat current at that time.

Premise alignment: The alignment is categorized as High—the premise aligned extremely highly for roughly a third of the target audience and aligned somewhat for the remainder of the department. Additionally, the whole of the targeted OU was on alert for any unpaid invoices following a recent event surrounding a legitimate unpaid invoice.

Target audience: One OU within NIST, handling financial matters (ordering and invoice reconciliation), administrative program support, and technical program support. $n = 73$

c) Order confirmation

Message description: The order confirmation phish appeared to be from, "Order Confirmation" (auto-confirm@discontcomputers.com). Note the misspelling of "discount" in the email address. The subject line was personalized and said, "[Firstname Lastname]Your order has been processed," with a space missing between the user's last name and the word "Your." At the top of the email was an image of several holiday packages, with the words, "Order Confirmation" in bold immediately below the holiday package image. There was no personalization in the body of the email,

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

nor was there a greeting of any type. The email body text said, "Thank you for ordering with us. Your order has been processed. We'll send a confirmation e-mail when your item ships." This was followed by the words, "Order Details" in orange with, "Order: #SGH-2548883-2619437" (the order number was in blue text). The next section of the email said, "Estimated Delivery Date: 12/02/2016" (the date was in green text), "Subtotal: $59.97," "Estimated Tax: $4.05," and "Order Total: $64.02" in bold. There was a large yellow button labeled with the text, "Manage order." The button was followed by the text, "Thank you for your order. We hope you return soon for more amazing deals." Near the bottom of the email was an image of a holiday snow globe and the text, "Need it in time for the holidays? Order before December 23 for free over-night shipping." ("December 23" was in blue). Much smaller gray text below that said, "Unless otherwise stated, items sold are subject to sales tax in in accordance with local laws. For more information, please view tax information." ("tax information" was in blue). Note the repeated word "in in," a subtle mistake that is very difficult for users to notice, especially given the small gray font. Finally, at the very bottom of the email appeared three additional links, all in blue on a single line: "Return Policy | Privacy | Account."

Premise alignment: The alignment is categorized as Medium—the premise aligned for those who had purchasing authority in the OU and for those who had recently placed an order, a small subset of the whole OU. However, the training exercise took place in December, when many people make on-line purchases for the holidays.

Target audience: One OU within NIST, handling financial matters (ordering and invoice reconciliation), administrative program support, and technical program support. $n = 66$

d) Gmail

Message description: The Gmail phish was a particularly clever spear phish. It targeted mid-level management using a spoofed upper management Gmail address, a tactic based on a real-world phish previously observed at NIST. It appeared to come from the personal Gmail account of NIST's director (firstname.lastname1@gmail.com) and went to a list of laboratory managers. The subject line was, "Safety Awareness," which is important given that NIST has a very strong emphasis on fostering a culture of safety. The email was personalized with the recipient's first name. The body said, "Please make sure your groups are aware of this new requirement:" with a link following this text. The email was signed simply with the first name of the organization's director.

Premise alignment: The alignment is categorized as High—the premise alignment is very strong given the larger organization's substantial emphasis on workplace safety and that the message appeared to come from NIST's director—a notable authority figure in this context. Alignment is further strengthened by the target department's responsibility for the larger organization's occupational health and safety.

Target audience: One OU within NIST, handling financial matters (ordering and invoice reconciliation), administrative program support, and technical program support. $n = 64$

e) Weblogs

Message description: The weblogs phish was another spear phish. It appeared to come from a system administrator with the email address, notice@nist.gov. The subject line was, "Unauthorized Web Site Access." There was no personalization. The body said, "*This is an automated email* Our regulators require we monitor and restrict certain website access due to content. The filter system flagged your computer as one that has viewed or logged into websites hosting restricted content. The system is not fool-proof and may incorrectly flag restricted content. The IT department does not investigate every web filter report, but disciplinary action may be taken." In bold, it said, "Log into the filter system with your network credentials immediately and review your logs to see which websites triggered this alert." This was followed by a link that was labeled, "Web Security Logs." There was no contact information given, and the email closed with, "Do not reply to this email. This email was automatically generated to inform you of a violation of our security and content policies."

Premise alignment: The alignment is categorized as High—the premise aligns with the fact that accessing inappropriate content is indeed a violation of the organization's Rules of Conduct policy and can be grounds for dismissal for anyone at the organization. The premise capitalizes on the fact that many organizations, including NIST, scan log data routinely. The threat component coupled with the severity of the consequences increases the alignment. Of note, all new employees receive in-person training regarding the organization's Rules of Conduct and IT policies, where the disciplinary actions associated with inappropriate web content viewing are highly stressed.

Target audience: One OU within NIST, handling financial matters (ordering and invoice reconciliation), administrative program support, and technical program support. $n = 73$

f) Valentine

Message description: The Valentine phish appeared to come from "eCard Delivery" with the email address, "do_not_reply@ecardalert.com." The subject line said, "Happy Valentine's Day! See who sent you an e-card…" There were three large red heart images at the top of the email. There was no personalization. The body of the email said, "A secret admirer wished you a Happy Valentine's Day! Some of you may have heard about our employee greeting cards that can be used to acknowledge fellow employees. Click on the link below to view yours." This was followed by a large link that said, "Your Card is Waiting," and additional text that said, "If you are having trouble viewing the e-card please click here." "Would you like to send an e-card? Visit our site. *Making someone's day, one e-card at a time…"* The email closed with, "This email may contain confidential and privileged information for the sole use of the intended recipient. If you are not the intended recipient, please contact the sender and delete all copies. Any review or distribution by others is strictly prohibited. Thank you." The Valentine phish was sent January 22, 2018, prior to Valentine's Day.

Premise alignment: The alignment is categorized as Low—the premise does not align with a business process but is more personal in nature. However, the message contains a sentence about using the service to acknowledge a fellow employee. So, while the premise does not align with a business practice, it does

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

play on the reader's curiosity, aligning with the upcoming occasion of Valentine's Day, of which most people are aware.

Target audience: All staff at NIST with an email address were targeted, from the human resources department, to finance, to bench scientists, to administrative support and all levels of management. $n = 4977$

g) Security token

Message description: The security token phish appeared to be from "Alerts" with the email address, "alerts@verifytoken.com." The subject line was, "Verify Your Security Token Was Not Compromised." The email was personalized using the format, "Lastname, Firstname, Middle Initial (Fed)." The body said, "Recently we have been made aware of a security breach in our security token product. Some of the tokens have been compromised and may need to be replaced. In order to find out if you [sic] token has been compromised, Validate Your Security Token Here." (Note the "you token" instead of "your token" here.) The email was signed "Rivest Shamir Adleman, Director of Identity and Access Management." The email closed with smaller text that said, "This email may contain confidential and privileged information for the sole use of the intended recipient. Any review or distribution by others is strictly prohibited. If you are not the intended recipient, please contact the sender and delete all copies. Thank you." If someone clicked the "Validate Your Security Token Here" link, they were taken to a data entry webpage, with the URL "secure.verifytoken.com." The top of the webpage said, Token Security with a red background, followed by "Attention!!! Recently the safety of some security tokens has been compromised. Enter your username and six-digit number that is generated every 60 seconds by your security token and we will know if you will need a new token. Should you need a new token, you will be given contact information to request a new token, which will be shipped to you overnight." This was followed by the text, "Account Login," with fields labeled, "User ID" and "Password or Passcode." There was a blue button labeled, "Login" and an "I'm not a robot" checkbox. At the bottom of the webpage was the text, "A passcode contains a PIN and a number from a security token."

Premise alignment: The alignment is categorized as Medium—the premise does not align at all for those personnel who do not have a security token (roughly 43 % of all staff at the organization). Further, the premise does not align for those staff who expect any token checking and replacement would be conducted via the organization rather than a third party—likely a significant portion of the remaining 57 % as the organization has a very strong posture regarding IT security.

Target audience: All staff at NIST with an email address were targeted, from the human resources department, to finance, to bench scientists, to administrative support and all levels of management. $n = 5024$

2) Determining difficulty ratings

a) Applying the Phish Scale

As described previously, the difficulty rating for an individual phishing message is determined first by categorizing the number of objectively observed cues and the premise alignment. Then use the conceptual framework in Table I to select the difficulty rating associated with the categorized number of cues and premise alignment. In Table III the Phish Scale ratings are shown for each of the seven previously described phishing exercises, including the number of cues for each email (detail provided in Appendix B), the premise alignment (from the exercise description), the difficulty rating (from the conceptual framework in Table I), and the actual click rates for each exercise.

The table in Appendix B contains the counts for each cue and a total count for each exercise. When counting cues in a given email message during analysis, it is important to note that these cue counts are based on our extremely careful scrutiny of the email messages; most email users are not going to notice or attend to all the available cues.

Note that in order to calculate the difficulty rating, the number of cues must be further categorized into *Few*, *Some*, or *Many*, in order of decreasing difficulty. Although some cues are more salient than others, we anticipate this is a reasonable first approximation. In this initial version of the Phish Scale, we propose the associated ranges as follows: the category labeled *Few* is represented by 1 to 8 cues, the category labeled *Some* by 9 to 14 cues, and the category labeled *Many* by 15 or more cues. These ranges are based on our existing dataset; at this stage of scale development, the click rates inform the categorization of the cue counts. We fully expect the cue count ranges may change with broader application of the Phish Scale to a larger variety of phishing emails. A larger corpus of phishing emails will be needed to validate the cue count ranges.

It should be emphasized that the number of cues alone does not determine the detection difficulty for a target audience; it is only when considered in conjunction with the premise alignment that a detection difficulty rating can be computed.

TABLE III. PHISHING EXERCISE DATA

Exercise	Number of cues	Premise alignment	Difficulty rating	Actual phishing click rate
New voicemail ($n = 69$)	11 (Some)	Medium	Moderately difficult	11.6 % (8/69)
Unpaid invoice ($n = 73$)	8 (Few)	High	Very difficult	20.5 % (15/73)
Order confirmation ($n = 66$)	18 (Many)	Medium	Moderately difficult	9.1 % (6/66)
Gmail ($n = 64$)	7 (Few)	High	Very difficult	49.3 % (39/73)
Weblogs ($n = 73$)	14 (Some)	High	Very difficult	43.8 % (28/64)
Valentine ($n = 4097$)	13 (Some)	Low	Moderately/ Least difficult	11.0 % (549/4977)
Security token ($n = 5024$)	12 (Some)	Medium	Moderately difficult	8.7 % (439/5024)

The security token phish was the only exercise with a data entry component—after clicking the link users were taken to a webpage requesting their credentials. We report the data entry rates here rather than in Table III. For the security token phish, 24.4 % (107/439) of clickers entered data on the credential-harvesting webpage. However, this is only 2.1 % (107/5024) of the total number of employees who received the phishing email. Given that roughly 75 % of clickers did not enter data on the

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

webpage, it seems that additional suspicion was triggered on this page, likely due to being asked for credentials. This is certainly in line with the fact that mandatory yearly security awareness training at NIST in the past has focused heavily on not sharing credentials.

b) Observations

Through these seven phishing exercises, we applied our Phish Scale to a variety of phishing attack types. This includes link-based attacks (New voicemail, Order confirmation, Gmail, Weblogs, and Valentine), an attachment attack (Unpaid invoice, which mimicked the real-world Locky ransomware attack), and a data entry or credential-harvesting attack (Security token). Now that we have used the Phish Scale to determine the detection difficulty rating for seven phishing exercises, there are a few observations we can make.

All of the exercises having a detection difficulty rating of *Very difficult* also have relatively high click rates (Unpaid invoice: 20.5 %, Gmail: 49.3 %, and Weblogs: 43.8 %). However, the Weblogs exercise has many more cues than the other two exercises, and at 14 cues, was at the extreme end of the *Some* cues range (9 to 14).

All of the exercises having a detection difficulty rating of *Moderately difficult* have relatively lower click rates (New voicemail: 11.6 %, Order confirmation: 9.1 %, and Security token: 8.7 %). The Valentine exercise has the detection difficulty range *Moderately difficult* to *Least difficult* and has a click rate of 11.0 %. The Valentine and Security token exercises have relatively larger and more varied sample sizes than the other exercises which likely makes it more difficult to categorize the premise alignment. And finally, we do not have an exercise with a detection difficulty rating of *Least difficult*, calling attention to the need to apply the Phish Scale to additional exercises.

c) Limitations

This work is an early effort to characterize phishing message detection difficulty for email users situated in their normal email processing environments. As such, the authors acknowledge there are certainly limitations with this work at this time.

Current notable limitations in this work include: 1) the list of cues is long but not exhaustive; 2) the uneven saliency of cues is not reflected; 3) categorizing premise alignment is not formulaic; 4) cue count ranges need to be informed by additional data; and, 5) additional data are needed for scale validation.

The authors anticipate that each of these limitations will be addressed as the Phish Scale is developed further.

V. Discussion and Future Directions

A. Click rates alone are insufficient: Why phishing detection difficulty matters

CISOs responsible for overseeing embedded phishing awareness training are often concerned when they observe click rates that are higher than expected. They are left wondering why click rates continue to be variable—possibly including large spikes—despite spending a significant amount of money and time training staff. CISOs must justify their cyber awareness training budgets and show a good ROI, lest their funding for such training be reduced. Unfortunately, if click rates continue to be high or variable, it is often—and we posit, incorrectly—perceived as due to ineffective training. We argue that this perception is fundamentally incorrect and hope to begin dispelling this perception through our development of a Phish Scale. Furthermore, we argue against focusing solely on phishing exercise click rates, and instead strongly encourage the inclusion of reporting rates and reporting times as well; these metrics must be considered in conjunction, not in isolation, as early reporting can greatly improve mitigation efforts. Are reporting rates higher than click rates? Is time to first report sooner than time to first click?

We hope to frame the discussion around high click rates in a way that makes sense to CISOs and argue that high click rates can indicate that users are being exposed to new, difficult, and contextually relevant phishing campaigns. We firmly believe difficult exercises actually improve user training effectiveness and awareness for real-world threats more than solely repeating the same or very similar, easier-to-detect phish. Click rates must be considered in conjunction with a deeper understanding of the phishing emails themselves and in light of reporting behavior as well. To this end, we have developed a Phish Scale to aid CISOs in better understanding and characterizing the detection difficulty of a given phishing exercise. Using operational data, the scale provides an indication of the difficulty email users in a target population will have detecting a particular phishing message. The Phish Scale addresses multiple components of phishing detection difficulty: cues, such as [26] and [30], and user context alignment [14]. Although our Phish Scale cue list is quite extensive, it is by no means exhaustive.

We expect that the three detection difficulty ratings we identified, *Very difficult*, *Moderately difficult*, and *Least difficult,* may eventually equate to click rate ranges. In speaking with CISOs, we anticipate ranges roughly along these lines: the *Very difficult* category having click rates above 20 %, the *Moderately difficult* category having click rates in the approximately 9 to 20 % range, and the *Least difficult* category having click rates below 9 %. We plan to inform the actual ranges with additional empirical data; the seven exercises presented here are a start.

It is early days for the Phish Scale, however, we believe the conceptual framework has promise when we consider the projected detection difficulty rating and the actual click rates for the seven exercises we examined. Additionally, we stress the Phish Scale components are still in development. We know all cues do not have equal salience. Finding an abbreviated method for CISOs to characterize premise alignment has proven difficult and elusive thus far. And finally, additional components such as severity of consequences and other factors may need to be considered sooner rather than later.

Indeed, we already see indications that additional factors warrant investigation for inclusion. For example, in the Weblogs exercise, there are 14 cues (categorized as *Some*), but this is right on the cusp of the *Many* cues category—and a corresponding easier detection difficulty rating. However, the actual click rate is very high, 43.8 % indicating detection is indeed difficult. The severity of the consequences in this premise is also very high—loss of job—suggesting its potentially strong influence.

B. Differential cue salience: Not all cues are created equally

Capturing the effect of phishing message cues is difficult, as not all cues are created equally. The saliency and effect of any

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

particular phishing cue varies, determining whether it is perceived as a suspicion indicator versus a compelling hook. This aspect of phishing message characteristics is important to note. Whether a cue is perceived as a phish indicator versus a hook depends on the user and the user's context when processing the email. Well-known phishing indicators such as misspellings and grammar errors are often regarded by email users as suspicion-generating, and when noticed can lead to additional user scrutiny of the message for more phishing indicators. Another undisputed phishing characteristic is urgency. Its use is so common that it should be a red flag, however, urgency is legitimately common-place in today's world, diluting its suspicion-generating signal strength. Additionally, urgency inhibits System 2 processing [22] making it more a hook enhancer than a red flag.

C. Categorizing user context and premise alignment

In this initial version of a Phish Scale, we have used the terms *High*, *Medium*, and *Low* to bucket premise alignment for a target audience into intuitive high-level categories with associated definitions. While this is sufficient for the beginning phase of scale development, we may seek to refine the characterization methods for these categorical variables in future work, by investigating contextual relevance measures and scales. "Contextual" is a part of existing scales in other domains such as, "A Contextual Measure of Achievement Motivation" [35] and "contextual performance" as a dimension of individual work performance [24]. How might such existing scales and measures be leveraged for use in the phishing domain? Additionally, how do we account for changes in context over time?

Changes in contextual relevance may occur over quite long timescales, as someone slowly adds or changes job responsibilities over the years of their career, or very short timescales, as some event that day/week/month may trigger heightened contextual relevance. For example, Greene et al. [14] explained that users were concerned over a real-world vendor invoice that was unpaid, leading to temporarily heightened contextual relevance for the unpaid invoice phishing email. Daily events, such as expecting or missing a phone call, can temporarily heighten the contextual relevance of a "new voicemail" phishing email. Factors such as being busy, stressed, or rushed can also fluctuate widely during a work day. It is likely the case that there is a relatively fixed component of user context, in addition to a more time-sensitive, variable component. The current Phish Scale does not break down context and associated premise alignment into these subcomponents. It is unclear whether such a fine-grained distinction is indeed necessary at this point.

Although it may be quite feasible to discern premise alignment with finer granularity than our existing categories, this may actually be superfluous for the intended audience of the Phish Scale. With our goal of developing a simple, easy to use Phish Scale for CISOs and those responsible for implementing and overseeing phishing awareness training programs, it is likely the case that *High*, *Medium*, and *Low* categories for premise alignment are sufficient. The important point we seek to emphasize with our Phish Scale is that a highly relevant context makes it extremely difficult for users to detect phishing emails. The greater the contextual relevance, the less likely a user is to notice, attend to, and think deeply about suspicious email cues. Daily stressors such as time pressure in general reduce the cognitive resources that users have available to dedicate to email processing. When cognitive resources are reduced, it makes it more likely that users will engage in faster, heuristic, System 1 processing rather than thoughtful, slower, deeper System 2 processing [22].

A final point with respect to user context and premise alignment has to do with the size of the target audience: categorizing premise alignment becomes more difficult as the size of the target audience increases. With a larger target audience, there is typically a much greater variety of work responsibilities present and a wider variety of user contexts, which may or may not align with a phishing email premise.

D. Comparing phishing data across sectors

Although cross-exercise and cross-sector phishing comparisons are frequently made, and are indeed quite valuable, interpretation of such comparisons still pose significant challenges. In particular, when the level of phishing detection difficulty can vary so dramatically based on user context and premise alignment, it is in some sense a meaningless comparison without a basic understanding and assessment of: 1) characteristics of the phishing email itself and 2) characteristics of the target user population. More specifically, one must understand the premise and cues contained within a given phish in conjunction with the work context of the target user population. Toward this end, we believe our Phish Scale shows great promise as a tool to help frame data sharing on click rates and reporting rates across exercises, organizations, and sectors.

As we refine and mature this tool with input from the larger usable security community, we hope to move the Phish Scale out of the research community and into operational use. For instance, we believe that beyond providing benefits to CISOs and phishing training implementers, our Phish Scale could also provide significant value to joint organizations responsible for sharing cyber threat intelligence data. For example, the Federal Bureau of Investigation (FBI), has an InfraGard program, a partnership between the FBI and the private sector dedicated to sharing information and intelligence [10]. There are other such collaborative programs as well, for example, the National Cyber-Forensics and Training Alliance (NCFTA) is a nonprofit partnership between private industry, government, and academia working together to disrupt cybercrime [27]. Phishing in particular, and social engineering in general, are active threats across all industry verticals. By providing a phishing difficulty rating framework, our Phish Scale can help facilitate collaboration using a common language surrounding human phishing threat detection.

E. Future work

We encourage other usable security researchers and practitioners to use our Phish Scale, apply it to a much wider variety of phishing emails, and test its predictions against both existing phishing training exercise data, and ultimately against real-world phishing emails as well. We plan to continue applying our Phish Scale to a larger corpus of additional emails for which we have click rate and premise alignment data, and plan to partner with external entities to do the same. Unfortunately, our access to concurrent reporting data is more limited. A notable challenge of conducting research with operational workplace data is that there is often a tradeoff

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

between experimental control and ecological validity. In this case, we had the benefit of extremely high ecological validity, as users were in their normal workplace settings with their normal tasks and email loads, but without the control necessary to capture reporting rates at the time of the phishing exercises. Nonetheless, the benefit of having new, in situ workplace data for ~ 5 000 employees offers an important contribution to the phishing literature and to the larger usable security community.

Beyond applying and testing the current Phish Scale with additional data, we intend to explore new scale components as well. For instance, it appears the Phish Scale would benefit from incorporating a measure of perceived consequence severity. Greene et al. [14] found that clickers were concerned over consequences arising from *not* clicking, such as failing to be responsive to their job duties. In contrast, non-clickers were more concerned over consequences due to clicking, such as accidentally downloading malware. Additionally, concern over consequences varied depending on the premise of the phish. In the phishing exercises described in the current paper, it is likely that concern over consequences was much higher for the Weblogs exercise, with its implied consequence of disciplinary action, to include dismissal, and the Security token exercise, which could have been concerning for teleworkers, as a security token is needed to access the organization's network from off campus. The current instantiation of the scale does not specifically address concern over consequences or the perceived relative severity of consequences. It may be possible that premise alignment alone is sufficient and already captures these effects for the Phish Scale's intended purpose, but additional research on this would be beneficial.

Additionally, we would like to investigate incorporating work on personality factors, and ultimately folding the various components of our Phish Scale into a lens model, an application of multiple regression often used in judgement and decision-making research. This would build upon prior lens modeling work by Tamborello and Greene [36] and Molinaro and Bolton [26]. Additional modeling and simulation research could explore the predicted click rates and reporting rates for different combinations of cues, context alignment, personality types, and phishing premises. How do different combinations affect phishing susceptibility? For example, consider this combination: users scoring high on conscientiousness, with a financial work context, who receive a phishing email with an authoritarian/time-sensitive transfer of funds premise, and very few suspicious cues. What if everything were the same but the work context, is that difference alone sufficient for someone to catch this phish? While we believe that context may trump all, additional research is necessary to see in which scenarios this holds, as well as, how and when it may change. One could simulate—with a well-validated model—the large number of possible combinations, to determine where to focus research and training intervention efforts based on quantified predicted risk metrics, such as the likelihood of clicking versus reporting.

F. Broader implications

In this section, we move beyond discussion of immediate future plans for the Phish Scale and into a discussion of broader implications for our work. The Phish Scale—and indeed phishing in general—is part of a much larger research agenda that addresses a spectrum of usable security issues. For instance, understanding risk, including human risk, is a key component of any organization's cybersecurity strategy, and risk management frameworks play an important role in helping maintain security and privacy [28]. Ultimately, we hope our Phish Scale can be used to help CISOs better understand and characterize their organization's phishing risk, by essentially profiling the types of phishing premises their users are more or less susceptible to as well as the organization's actual threats. Such data can be used to prioritize training efforts on more targeted interventions, and to prioritize investigative efforts for real-world suspected phishes. Targeted training interventions will likely need to move beyond embedded phishing exercises, especially for repeat clickers. In-person seminars, posters, informal lunch and learn sessions, and so on, are all part of a larger security awareness program. Additional interventions may include special email Graphical User Interface (GUI) elements or flagging, or perhaps more aggressive email filtering for certain users or groups based on their risks and job responsibilities.

In addition to risk profiling and targeted training, future work is also needed to understand how new technological email security measures will impact phishing. In particular, government agencies are quickly moving toward email authentication by implementing protocols such as Domain-based Message Authentication, Reporting, and Conformance (DMARC) and Domain Keys Identified Mail (DKIM) per the Department of Homeland Security (DHS) Binding Operational Directive 18-01 [6]. How will that affect the phishing space? On the other hand, pretexting is already gaining in popularity and will likely continue to do so, especially if new technological solutions prevent or threaten the success of certain more "traditional" phishing email scams. As advances in technological protections make some attacks less effective, or even one day obsolete, the attacks will not stop, but rather will transition and evolve in response. For instance, it seems likely that other out-of-band social engineering methods will continue to gain in popularity. Phishing is but one component of a much larger social engineering problem facing the cybersecurity field. Future work should examine how lessons learned in the phishing domain may inform other varieties of social engineering problems as well.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the institution's Information Technology Security and Networking Division and our Information Technology Security Officer partner for their support throughout this project, and thank the CISOs and training implementers who spoke with us regarding their phishing awareness programs, as well as anonymous reviewers for their comments.

DISCLAIMER

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

REFERENCES

[1] M. Alsharnouby, F. Alaca, and S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks." International Journal of Human-Computer Studies, 2015, 82, pp. 69–82.

[2] M. Blythe, H. Petrie, and J.A. Clark, "F for fake: Four studies on how we fall for phish," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM. May 2011, pp. 3469–3478.

[3] C.I. Canfield, B. Fischhoff, A. Davis, "Quantifying phishing susceptibility for detection and behavior decisions." Human Factors: The Journal of the Human Factors and Ergonomics Society, 2016, 58, pp. 1158–1172.

[4] D. Caputo, S.L. Pfleeger, J. Freeman, and M. Johnson, "Going spear phishing: Exploring embedded training and awareness," in IEEE Security & Privacy, 12(1), January 2014, pp. 28–38.

[5] Cybersecurity Ventures, "2017 Cybercrime Report," https://1c7fab3im83f5gqiow2qqs2k-wpengine.netdna-ssl.com/2015-wp/wp-content/uploads/2017/10/2017-Cybercrime-Report.pdf (Accessed Nov 2018).

[6] DHS, Department of Homeland Security. Binding Operational Directive BOD-18-01, 2017, https://cyber.dhs.gov/assets/report/bod-18-01.pdf (Accessed Nov 2018).

[7] R. Dhamija, J.D. Tygar, M. Hearst, "Why phishing works," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2006, pp. 581–590.

[8] J.S. Downs, M. Holbrook, and L.F. Cranor, "Decision strategies and susceptibility to phishing," in Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS '06), ACM, 2006, pp. 79–90.

[9] S. Egelman, L.F. Cranor, and J. Hong, "You've been warned: An empirical study of the effectiveness of web browser phishing warnings," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2008, pp. 1065–1074.

[10] FBI, Federal Bureau of Investigations. https://www.fbi.gov/about/partnerships/infragard (Accessed Nov 2018)

[11] B.J. Fogg, "Prominence-interpretation theory: Explaining how people assess credibility online." In CHI'03 Extended Abstracts on Human Factors in Computing Systems, ACM, 2003, pp. 722–723.

[12] S. Furnell, "Phishing: can we spot the signs?" Computer Fraud and Security 2007, pp. 10–15.

[13] S. Grazioli, "Where did they go wrong? An analysis of the failure of knowledgeable internet consumers to detect deception over the internet." Group Decision and Negotiation, 2004, 13, pp. 149–172.

[14] K.K. Greene, M. Steves, M. Theofanos, and J. Kostick, "User Context: An Explanatory Variable in Phishing Susceptibility." USEC NDSS 2018. Usable Security Workshop at the Network and Distributed Systems Security Symposium. February 18, 2018. San Diego, CA. DOI: https://dx.doi.org/10.14722/usec.2018.23016

[15] K.K. Greene, M. Steves, and M. Theofanos. "No Phishing beyond This Point," in IEEE Computer, Cybertrust Column, 2018, vol. 51, no. 6, pp. 86–89. DOI: http://doi.ieeecomputersociety.org/10.1109/MC.2018.2701632

[16] Y. Han and Y. Shen. "Accurate spear phishing campaign attribution and early detection." In Proceedings of the 31st Annual ACM Symposium on Applied Computing, ACM, 2016, pp. 2079–2086.

[17] Healthcare CyberGard Annual Conference, Charlotte, October 2018, https://www.ncfta.net/healthcare-cybergard-annual-conference-charlotte-october-2018/ (Accessed Jan 2019).

[18] Information Security and Privacy Advisory Board, https://csrc.nist.gov/CSRC/media/Projects/ISPAB/documents/minutes/ispab-june-2018-meeting-minutes.pdf (Accessed Jan 2019).

[19] M. Jakobsson, "The human factor in phishing." Privacy and Security of Consumer Information, 2007, 7, pp. 1–19.

[20] M. Jakobsson and P. Finn, "Designing and conducting phishing experiments." IEEE Technology and Society Magazine, Special Issue on Usability and Security, 2007.

[21] A. Karakasiliotis, S. Furnell, and M. Papadaki, "Assessing end-user awareness of social engineering and phishing." In Australian Information Warfare and Security Conference, School of Computer and Information Science, 2006, Edith Cowan University, Perth, Western Australia.

[22] D. Kahneman, "Thinking, Fast and Slow." New York: Farrar, Straus and Giroux, 2011.

[23] D. Kim, and J. Hyun Kim, "Understanding persuasive elements in phishing e-mails: A categorical content and semantic network analysis", Online Information Review, 2013, Vol. 37 Issue: 6, pp. 835–850, https://doi.org/10.1108/OIR-03-2012-0037

[24] L. Koopmans, C.M. Bernaards, V.H. Hildebrandt, H.C. de Vet, and A.J. van der Beek, "Measuring individual work performance: Identifying and selecting indicators." *Work*, 2014, 48(2), pp. 229–238.

[25] D.E. Levari, D.T. Gilbert, T.D. Wilson, B. Sievers, D.M. Amodio, and T. Wheatley, "Prevalence-induced concept change in human judgement," *Science* 29, June 2018: Vol. 360, Issue 6396, pp. 1465–1467. DOI: https://doi.org/10.1126/science.aap8731

[26] K.A. Molinaro, and M.L. Bolton, "Evaluating the applicability of the double system lens model to the analysis of phishing email judgments." *Computers & Security*, 2018, 77, pp. 128–137.

[27] NCFTA, National Cyber-Forensics and Training Alliance. https://www.ncfta.net (Accessed Nov 2018).

[28] NIST, National Institute of Standards and Technology, 2018, "Special Publication 800-37, Revision 2, Risk Management Framework for Information Systems and Organizations—A System Life Cycle Approach for Security and Privacy"

[29] L. H. Newman, "What Spammers Could Do with Your Hacked Facebook Data," in Wired, 2018, https://www.wired.com/story/facebook-hack-data-spammers/ (Accessed Nov 2018).

[30] K. Parsons, M. Butavicius, M. Pattinson, A. McCormac, D. Calic, and C. Jerram, "Do Users Focus on the Correct Cues to Differentiate Between Phishing and Genuine Emails?", Australasian Conference on Information Systems 2015, https://arxiv.org/abs/1605.04717 (accessed Nov 2018).

[31] K. Parsons, A. McCormac, M. Pattinson, M. Butavicius, C. Jerram, "Phishing for the truth: A scenario-based experiment of users behavioural response to emails." In IFIP International Information Security Conference, Springer, Berlin, Heidelberg, 2013, pp. 366–378.

[32] Ransomware, https://www.trendmicro.com/vinfo/us/security/definition/RANSOMWARE (Accessed Jan 2019).

[33] R. W. Rogers, "A protection motivation theory of fear appeals and attitude change." J. Psychol, 1975, 91, pp. 93–114.

[34] B. D., Sawyer and P.A. Hancock, "Hacking the Human: The Prevalence Paradox in Cybersecurity." Human Factors, 2018, 60(5), pp. 597–609.

[35] R.L. Smith, "A contextual measure of achievement motivation: Significance for research in counseling." *VISITAS Online*, 2015.

[36] F.P. Tamborello, K.K. Greene. "Exploratory Lens Model of Decision-Making in a Potential Phishing Attack Scenario." National Institute of Standards and Technology Interagency Report, NISTIR 8194, October 2017, DOI: https://doi.org/10.6028/NIST.IR.8194

[37] A. Tsow, M. Jakobsson, "Deceit and deception: A large user study of phishing." Indiana University, Technical report TR649, 2007.

[38] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H.R. Rao. "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model." *Decision Support Systems*, vol. 51 no. 3, March 2011, pp. 576–586.

[39] A. Vishwanath, B. Harrison, Y.J. Ng, "Suspicion, cognition, and automaticity model of phishing susceptibility." Communication Research, 2016, 0093650215627483.

[40] J. Wang, T. Herath, R. Chen, A. Vishwanath, and H.R. Rao, "Phishing susceptibility: An investigation into the processing of a targeted spear phishing email," in IEEE Transactions on Professional Communication, vol. 55, no. 4, December 2012, pp. 345–362.

[41] J. Wang, Y. Li, R. Rao, "Coping responses in phishing detection: an investigation of antecedents and consequences." Inf. Syst. Res., 2017, 28, pp. 378–396.

[42] E.J. Williams, J. Hinds, and A.N. Joinson, "Exploring susceptibility to phishing in the workplace." *International Journal of Human-Computer Studies*, 2018, 120, pp. 1–13.

[43] R. Wright, S. Chakraborty, A. Basoglu, and K. Marett, "Where did they go right?" Understanding the deception in phishing communications. Group Decision and Negotiation, 2010, 19, pp. 391–416.

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

APPENDIX A

The table below contains the list of operationalized cues, indicators and hooks we used to count the phishing message characteristics to obtain a rating of phishing detection difficulty. For each cue, the associated references and the criteria we used for counting are also provided.

TABLE: OPERATIONALIZED CUES

Cue Type	Cue Name	Description	References	Criteria for Counting
Error	Spelling and grammar irregularities	Spelling or grammar errors, mismatched plurality and so on	[2], [3], [8], [12], [13], [19], [20], [21], [31], [38], [40], [43]	Does the message contain spelling or grammar errors, including mismatched plurality?
	Inconsistency	Inconsistent content within the email	[14]	Are there inconsistencies contained in the email message?
Technical indicator	Attachment type	The presence of file attachments, especially an executable	[16]	Is there a potentially dangerous attachment?
	Sender display name and email address	Spoofed display names - hides the sender and reply-to email addresses	[2], [8], [12], [21], [23], [38], [39], [40]	Does a display name hide the real sender?
	URL hyperlinking	URL hyperlinking hides the true URL behind text; the text can also look like another link	[3], [8], [9], [12], [19], [21]	Is there text that hides the true URL in a hyperlink?
	Domain spoofing	Domain name used in email address and links looks similar to plausible	[14], [37]	Is a domain name used in addresses or links plausibly similar to a legitimate entity's domain?
Visual presentation indicator	No/minimal branding and logos	No or minimal branding and logos	[2], [12], [13], [19], [21], [23], [37], [39]	Is appropriate branding missing?
	Logo imitation or out-of-date branding/logos	Spoof or imitation of logo/out-of-date logo	[14]	Do any branding elements appear to be an imitation or out-of-date?
	Unprofessional looking design or formatting	Formatting and design elements that do not appear to have been professionally generated	[7], [11], [19], [21], [31], [37]	Does the design and formatting violate any conventional professional practices?
	Security indicators and icons	Security indicators and icons	[7], [19]	Are any inappropriate security indicators or icons present?
Language and content	Legal language/copyright info/disclaimers	Any legal type language such as copyright information, disclaimers, tax implications	[19]	Does the message contain any legal type language such as copyright information, disclaimers, tax information?
	Distracting detail	Distracting Detail	[14]	Does the message contain any detailed aspects that are not central to the content?
	Requests for sensitive information	Requests for sensitive information, like a Social Security number or other identifying information	[8], [12], [14]	Does the message contain a request for any sensitive information, including personally identifying information or credentials?
	Sense of urgency	Use of time pressure to try to get users to quickly comply with the request	[1], [3], [8], [12], [21], [23], [38]	Does the message contain time pressure, including implied?
	Threatening language	Use of threats such as legal ramifications	[1], [3], [8], [21], [23], [38]	Does the message contain a threat, including an implied threat?
	Generic greeting	A generic greeting and an overall lack of personalization in the email	[1], [3], [8], [9], [21], [31], [37]	Does the message lack a greeting or lack personalization in the message?
	Lack of signer details	Emails including few details about the sender, such as contact information	[23]	Does the message lack detail about the sender, such as contact information?
Common tactic	Humanitarian appeals	Appeals to help others in need	[21], [23]	Does the message make an appeal to help others?

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

	Too good to be true offers	Contest winnings or other unlikely monetary and/or material offerings	[13], [21], [31], [43]	Does the message offer anything that is too good to be true, such having won a contest, lottery, free vacation and so on?
	You're special	Just for you offering… such as a valentine e-card from a secret admirer		Does the message offer anything just for you?
	Limited time offer	This offer won't last long…		Does the message offer anything for a limited time?
	Mimics a work or business process	Mimics any plausible work process such as new voicemail, package delivery, order confirmation, notice of invoice, and so on		Does the message appear to be a work or business-related process?
	Poses as friend, colleague, supervisor, authority figure	Email purporting to be from a friend, colleague, boss or other authority figure	[42]	Does the message appear to be from a friend, colleague, boss or other authority entity?

APPENDIX B

The table below contains the observed counts of each cue for each exercise we evaluated for this effort to-date.

TABLE: EXERCISE CUE COUNTS

		Exercise Cues Observed with Counts						
		11	8	18	7	14	13	12
Cue Type	***Cue Name***	**New Voicemail**	**Unpaid Invoice**	**Order Confirmation**	**Gmail**	**Weblogs**	**Valentine**	**Token**
Error	Spelling and grammar irregularities	1	1	2				1
	Inconsistency		1					1
Technical indicator	Attachment type		1					
	Sender display name and email address	1	1	1	1	1	1	1
	URL hyperlinking	1		6	1	1	3	1
	Domain spoofing	1	1	1	1	1	1	1
Visual presentation indicator	No/minimal branding and logos	1		1			1	1
	Logo imitation or out-of-date branding/logos							
	Unprofessional design or formatting	1					1	
	Security indicators and icons							
Language and content	Legal language/copyright info/disclaimers	1		1		1	1	1
	Distracting detail	2		2		1	1	
	Requests for sensitive information					1		1
	Sense of urgency		1	1	1	1	1	1
	Threatening language					3		
	Generic greeting			1	1	1	1	1
	Lack of signer details	1		1		1	1	1
Common tactic	Humanitarian appeals							
	Too good to be true offers							

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

	You're special						1	
	Limited time offer							
	Mimics a work or business process	1	1	1	1	1		1
	Poses as friend, colleague, supervisor, authority figure		1		1	1		

Steves, Michelle; Greene, Kristen; Theofanos, Mary.
"A Phish Scale: Rating Human Phishing Message Detection Difficulty."
Paper presented at 2019 Workshop on Usable Security and Privacy (USEC), San Diego, CA, United States. February 24, 2019 - February 24, 2019.

Metrology Requirements for Next Generation of Semiconductor Devices

George Orji
Microsystems and Nanotechnology Division, Physical Measurement Laboratory, NIST, Gaithersburg MD 20899

INTRODUCTION

The first International Roadmap for Devices and Systems (IRDS)[1] was published in 2018 and builds on the decades-long effort by the International Technology Roadmap Semiconductors (ITRS). The IRDS Metrology Chapter identifies emerging measurement challenges from devices, systems, and integration in the semiconductor industry and describes research and development pathways for meeting them, covering the next 15 years[1]. This includes, but is not limited to, measurement needs for extending CMOS, accelerating beyond CMOS technologies, novel communication devices, sensors and transducers, materials characterization and structure-function relationships.

Although devices based on traditional CMOS architectures are expected to reach their physical limits in the next few years, the devices and materials involved are more complex and difficult to measure than ever before[2]. The nanoscale sizes mean that the same fundamental limitations that will affect device performance also affect available metrology methods[3]. In addition to nanoscale size and complex structure, next generation devices will incorporate new materials such as graphene and transition metal dichalcogenide films. Because of changes in materials properties, measurements of film thickness and other parameters will require considerably more information about the layer-dependent material properties. This could be challenging for existing metrology techniques. The presentation will outline some of the key materials and lithography metrology challenges and highlight promising new techniques in an era of not only increased complexity, but one where scaling is no longer the main industry driver.

DEVICE AND LITHOGRAPHY OPTIONS

With the proliferation of non-planar device architectures, a key challenge for metrologists has been to develop the techniques required to obtain full three-dimensional device structure information. The introduction of gate all around (GAA) structures (lateral GAA and vertical GAA) and monolithic 3D structures[4] would make this even more challenging. Some of the challenges of GAA include small target volumes, localized information, and low signal to noise ratios. In addition to the above issues, monolithic 3D has the problem of non-uniform sensitivities at different depths. This means that metrology solutions would need to have a large depth of focus or be transmissive. In addition, 3D stacked chips and 3D very large-scale integration (3D VLSI) are fully functional tiers, so destructive characterization would be prohibitively expensive. Table 1 lists some of the device and lithography metrology challenges.

Beyond classical CMOS, most of the proposed device candidates, such as 1D-2D field effect transistors, lateral and vertical heterostructures, include the use of 2D materials (such as graphene and molybdenum disulfide), which are susceptible to beam damage. In addition to complex device structures, specific lithography options have their own challenges; for example, extreme ultra-violet (EUV) lithography has problems with mask defectivity, line-edge roughness and stochastics. Nanoimprint lithography's metrology challenges include defectivity, overlay, and template inspection. In addition to defect inspection, directed self-assembly has unique challenges with overlay and defectivity.

Preprint

Orji, Ndubuisi.
"Metrology requirements for next generation of semiconductor devices."
Paper presented at 2019 International Conference on Frontiers of Characterization and Metrology for Nanoelectronics (FCMN) , Monterrey, CA, United States. April 2, 2019 - April 4, 2019.

TABLE 1. Device structure and lithography metrology challenges

	Lateral Gate All Around	Vertical Gate All Around	3DVLSI/ Monolithic 3D
Patterning	193i, EUV	193i, EUV	193i, High NA, EUV+(DSA)
Channel Material	Ge, IIV(TFET)	Ge, IIV(TFET)	Ge, IIV(TFET) 2D Materials
Metrology Challenges (Select Examples)	• Small target volumes. • Localized information • Low SNR. • Non-uniform sensitivities	• Small target volumes. • Localized information • Low SNR. • Non-uniform sensitivities • Future metrology techniques may be destructive.	• Different materials on multiple levels. • Low contrast materials • Small target volumes. • Localized information • Low SNR. • Non-uniform sensitivities at different depths. • Nonsystematic DSA overlay shift. • Potential beam damage. • Low image contrast • Reduced cross-scattering due to small sizes • Difficulty obtaining optical properties (*n* & *k*)

DSA, directed self-assembly; TFET, tunnel field-effect transistor; SNR, signal-to-noise ratio; NA, numerical aperture

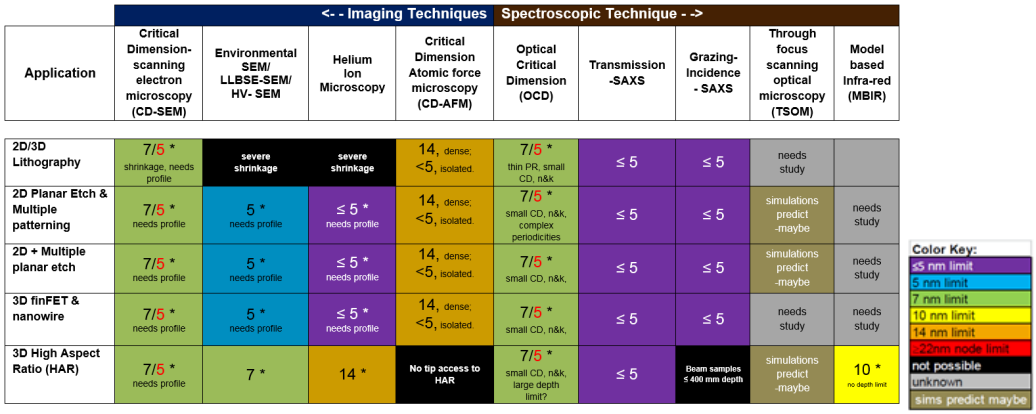


FIGURE 1: Lithography metrology gaps and limits. Continuous improvement and combined use of multiple methods could extend the applicability of some of these techniques[5-8]. All values are in nanometers; color key refers to limits of measurement techniques; LLBSE, low loss back-scattered electrons; SAXS, small angle x-ray scattering, HV, high voltage; Figure courtesy of B. Bunday[9].

POSSIBLE METROLOGY SOLUTIONS

Progress has been made in addressing many of the challenges listed in Table 1, but there continues to be the need for an additional broad range of metrology solutions commensurate with the complexities of the problems[5, 10]. Figure 1 shows metrology capabilities and approximate size limit needs for a wide range of lithography applications. The range of measurements needed to characterize different aspects of 3D features means that a wide variety of tools and instruments are required[11]. No single technique has the needed resolution, range, and low levels of uncertainty required to enable it to fully characterize these features.

A metrology approach that is gaining wider application is hybrid metrology, which relies on the complimentary use of multiple instruments. Figure 2 shows a conceptual diagram of multiple instruments being used to characterize a device. Each technique shown (scanning electron microscopy[2, 6], atomic force microscopy[12, 13], critical dimension x-ray scattering[14], scatterometry[15], and transmission electron microscopy[16, 17]) provides a specific capability[18] that the others do not have. In addition to multiple instruments, hybrid metrology also includes the use of statistical and combinatory techniques[19] that allow complementary analysis of the same features using the best measurement attributes of each technique.

Other promising methods include the use of ptychography-based methods to enhance electron, optical and X-ray based methods. Electron ptychography techniques were recently demonstrated for imaging 2D materials without causing beam damage, achieving a resolution of 0.04 nm[20]. At a larger length scale, X-ray ptychography methods were recently used to image and reconstruct whole chips with a resolution of 14.6 nm over a 10 μm range[21].

Orji, Ndubuisi.
"Metrology requirements for next generation of semiconductor devices."
Paper presented at 2019 International Conference on Frontiers of Characterization and Metrology for Nanoelectronics (FCMN) , Monterrey, CA, United States. April 2, 2019 - April 4, 2019.

Machine learning and other advanced analytics[22] techniques are gaining wide application in metrology[2]. This goes beyond data analysis and classification, and extends to instrument and measurement process optimization, including hybrid and virtual metrology.

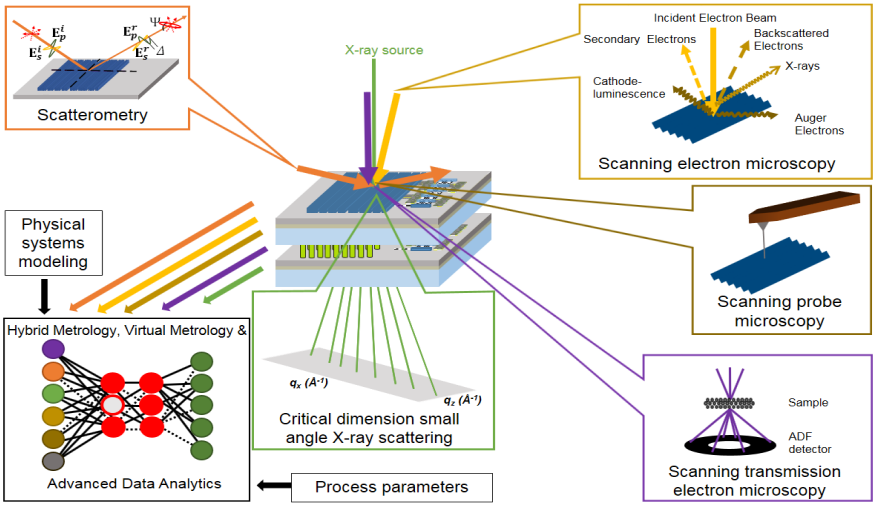


FIGURE 2: Multi-Instrument evaluation of a 3D stacked chip. Increasingly, advanced data analytics plays an increasingly major role in synthesizing information from multiple instruments[23-30] and process parameters. Figure courtesy of G. Orji and B. Barnes[2].

REFERENCES

1. *International Roadmap for Devices and Systems (IRDS) Metrology Chapter*, 2017 ed., Piscataway, NJ: IEEE 2018.

2. N. G. Orji, et al., *Metrology for the next generation of semiconductor devices, Nature Electronics* **1** (10), 532-547 (2018). https://doi.org/10.1038/s41928-018-0150-9

3. Z. Ma and D. G. Seiler, *Metrology and Diagnostic Techniques for Nanoelectronics*. New York: Pan Sanford 2017.

4. E. P. DeBenedictis, et al., *Sustaining Moore's law with 3D chips, Computer* **50** (8), 69-73 (2017). https://doi.org/10.1109/MC.2017.3001236

5. J. A. Liddle, et al., *Electron beam-based metrology after CMOS, APL Materials* **6** (7), 070701 (2018). https://doi.org/10.1063/1.5038249

6. J. S. Villarrubia, et al., *Scanning electron microscope measurement of width and shape of 10nm patterned lines using a JMONSEL-modeled library, Ultramicroscopy* **154**, 15-28 (2015). https://doi.org/10.1016/j.ultramic.2015.01.004

7. R. Attota and R. G. Dixson, *Resolving three-dimensional shape of sub-50 nm wide lines with nanometer-scale sensitivity using conventional optical microscopes, Appl. Phys. Lett.* **105** (4), 043101 (2014). https://doi.org/10.1063/1.4891676

8. J. Choi, et al., *Evaluation of carbon nanotube probes in critical dimension atomic force microscopes, J. Micro/Nanolith. MEMS MOEMS* **15** (3), 034005 (2016). https://doi.org/10.1117/1.Jmm.15.3.034005

9. B. D. Bunday, et al., *7/5nm logic manufacturing capabilities and requirements of metrology*, in *SPIE 10585*, 105850I. (2018). https://doi.org/10.1117/12.2296679

10. A. Masafumi, et al., *Metrology and inspection required for next generation lithography, Jpn J Appl Phys* **56** (6S1), 06GA01 (2017). https://doi.org/10.7567/JJAP.56.06GA01

11. N. G. Orji, et al., *Strategies for nanoscale contour metrology using critical dimension atomic force microscopy*, in *Instrumentation, Metrology, and Standards for Nanomanufacturing, Optics, and Semiconductors V*, 8105, p. 810505: SPIE. (2011). https://doi.org/10.1117/12.894416

12. N. G. Orji, et al., *Progress on implementation of a CD-AFM-based reference measurement system*, in *Proc. SPIE 6152*, 61520O. (2006). https://doi.org/10.1117/12.653287

13. R. Dixson, et al., *Multilaboratory comparison of traceable atomic force microscope measurements of a 70-nm grating pitch standard, J. Micro/Nanolith. MEMS MOEMS* **10** (1), 013015 (2011). https://doi.org/10.1117/1.3549914

14. R. J. Kline, et al., *X-ray scattering critical dimensional metrology using a compact x-ray source for next generation semiconductor devices, J. Micro/Nanolith. MEMS MOEMS* **16** (1), 014001 (2017). https://doi.org/10.1117/1.jmm.16.1.014001

15. A. C. Diebold, et al., *Perspective: Optical measurement of feature dimensions and shapes by scatterometry, APL Materials* **6** (5), 058201 (2018). https://doi.org/10.1063/1.5018310

16. N. G. Orji, et al., *Transmission electron microscope calibration methods for critical dimension standards, J. Micro/Nanolith. MEMS MOEMS* **15** (4), 044002 (2016). https://doi.org/10.1117/1.Jmm.15.4.044002

17. N. G. Orji, et al., *Tip characterization method using multi-feature characterizer for CD-AFM, Ultramicroscopy* **162**, 25-34 (2016). https://doi.org/10.1016/j.ultramic.2015.12.003

18. R. Dixson, et al., *Interactions of higher order tip effects in critical dimension-AFM linewidth metrology, J. of Vac. Sci. & Technol. B* **33** (3), 031806 (2015). https://doi.org/10.1116/1.4919090

19. N. F. Zhang, et al., *Improving optical measurement uncertainty with combined multitool metrology using a Bayesian approach, Appl. Opt.* **51** (25), 6196 (2012). https://doi.org/10.1364/ao.51.006196

20. Y. Jiang, et al., *Electron ptychography of 2D materials to deep sub-ångström resolution, Nature* **559** (7714), 343-349 (2018). https://doi.org/10.1038/s41586-018-0298-5

21. M. Holler, et al., *High-resolution non-destructive three-dimensional imaging of integrated circuits, Nature* **543** (7645), 402 (2017). https://doi.org/10.1038/nature21698

22. N. Rana, et al., *Leveraging advanced data analytics, machine learning, and metrology models to enable critical dimension metrology solutions for advanced integrated circuit nodes, J. Micro/Nanolith. MEMS MOEMS* **13** (4), 041415 (2014). https://doi.org/10.1117/1.jmm.13.4.041415

23. M. Kuhn, et al., *Transistor Strain Measurement Techniques and Their Applications*, in *Metrology and Diagnostic Techniques for Nanoelectronics*, Ed. by Z. Ma and D. G. Seiler New York: Pan Stanford, 2017, pp. 207-376. (2017).

24. S. O'Mullane, et al., *Advancements in Ellipsometric and Scatterometric Analysis*, in *Metrology and Diagnostic Techniques for Nanoelectronics*, Ed. by Z. Ma and D. G. Seiler New York: Pan Stanford, 2017, pp. 65-108. (2017).

25. N. G. Orji and R. G. Dixson, *3D-AFM Measurements for Semiconductor Structures and Devices*, in *Metrology and Diagnostic Techniques for Nanoelectronics*, Ed. by Z. Ma and D. G. Seiler New York: Pan Stanford, 2017, pp. 109-152. (2017).

26. N. G. Orji, et al., *Contour metrology using critical dimension atomic force microscopy, J. Micro/Nanolith. MEMS MOEMS* **15** (4), 044006 (2016). https://doi.org/10.1117/1.jmm.15.4.044006

27. C. McGray, et al., *Robust auto-alignment technique for orientation-dependent etching of nanostructures, J. of Micro/Nanolithography, MEMS, and MOEMS* **11** (2), 1-7 (2012). https://doi.org/10.1117/1.JMM.11.2.023005

28. D. Sunday and R. Kline, *X-Ray Metrology for Semiconductor Fabrication*, in *Metrology and Diagnostic Techniques for Nanoelectronics*, Ed. by Z. Ma and D. G. Seiler New York: Pan Stanford, 2017, pp. 31-64. (2017).

29. A. Vladár, *Model-Based Scanning Electron Microscopy Critical-Dimension Metrology for 3D Nanostructures*, in *Metrology and Diagnostic Techniques for Nanoelectronics*, Ed. by Z. Ma and D. G. Seiler New York: Pan Stanford, 2017, pp. 3-30. (2017).

30. V. Vartanian, et al., *TSV reveal height and dimension metrology by the TSOM method*, in *Metrology, Inspection, and Process Control for Microlithography XXVII*, 8681, p. 86812F. (2013). https://doi.org/10.1117/12.2012609

KEYWORDS

Semiconductor metrology, gate all around. 3D VLSI, nanometrology

Motivating Cybersecurity Advocates: Implications for Recruitment and Retention

Julie M. Haney
julie.haney@nist.gov
National Institute of Standards and Technology &
University of Maryland, Baltimore County
Gaithersburg, Maryland

Wayne G. Lutters
lutters@umd.edu
University of Maryland
College Park, Maryland

ABSTRACT

Given modern society's dependence on technological infrastructure vulnerable to cyber-attacks, the need to expedite cybersecurity adoption is paramount. Cybersecurity advocates are a subset of security professionals who promote, educate about, and motivate adoption of security best practices and technologies as a major component of their jobs. Successfully recruiting and retaining advocates is of utmost importance. Accomplishing this requires an understanding of advocates' motivations and incentives and how these may differ from other cybersecurity professionals. As the first study of its kind, we interviewed 28 cybersecurity advocates to learn about their work motivations. Findings revealed several drivers for cybersecurity advocacy work, most of which were intrinsic motivators. Motivations included interest in the field, sense of duty, self-efficacy, evidence of impact, comradery, and, to a lesser degree, awards and monetary compensation. We leverage these insights for recommendations on how to frame cybersecurity advocacy as a profession that fuels these motivations and how to maintain this across advocates' careers.

CCS CONCEPTS

• **Social and professional topics** → **Computing occupations**;
• **Security and privacy** → *Social aspects of security and privacy*.

KEYWORDS

Cybersecurity; advocacy; motivations; recruitment; retention

ACM Reference Format:
Julie M. Haney and Wayne G. Lutters. 2019. Motivating Cybersecurity Advocates: Implications for Recruitment and Retention. In *SIGMIS-CPR '19: ACM SIGMIS Computers and Personnel Research, June 20–22, 2019, Nashville, TN*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3322385.3322388

1 INTRODUCTION

Even with a rise in the frequency and severity of cyber-attacks, people often fail to adequately implement security best practices and technologies [28]. Given modern society's dependence on technology, the need for implementing effective cybersecurity ("the activity or process, ability or capability, or state whereby information and communications systems and the information contained therein are protected from and/or defended against damage, unauthorized use or modification, or exploitation" [21]) is paramount. A critical role in this adoption is the cybersecurity advocate who, recognizing that technology alone cannot solve security problems, is adept at addressing the interpersonal, societal, economic, and organizational factors often impeding adoption.

Cybersecurity advocates are security professionals who promote, educate about, and motivate adoption of security best practices and technologies as a major component of their jobs. In addition to technical knowledge of the security domain, this role requires non-technical competencies, such as interpersonal skills, communication skills, and context awareness [16]. Advocates' audiences are diverse and may include home users, office workers, students, technical staff, developers, and executives. Examples of cybersecurity advocates include security awareness professionals, secure development champions, those who advocate for security frameworks, and security consultants.

Due to the emphasis on technical skills within the cybersecurity field [10] and an estimated worldwide cybersecurity workforce shortfall of three million [17], there may be a dearth of professionals possessing the mix of technical and non-technical skills required for cybersecurity advocacy. Therefore, recruiting new advocates and retaining those already in the role is of utmost importance. Accomplishing this requires an understanding of advocates' motivations and incentives and how these may differ from other cybersecurity professionals.

To address this gap, we conducted interviews of 28 cybersecurity advocates. This paper reports on a subset of results from our first-of-its-kind investigation of cybersecurity advocates' work practices. Our previous papers focused on cybersecurity advocate skills and characteristics [16] and how advocates overcome people's negative perceptions of security [15]. In this paper, we extrapolate implications for recruitment and retention by analyzing answers to the following research questions from the broader study:

- What are the motivations of cybersecurity advocates?
- What is most rewarding about their advocacy jobs?

Our findings revealed a number of drivers for cybersecurity advocacy work, most of which were intrinsic motivators. Motivations included interest in the field, sense of duty, self-efficacy, evidence of impact, comradery, and, to a lesser degree, awards and monetary compensation. In particular, sense of duty and evidence of impact are intrinsically tied to advocates' roles as change agents and educators and their direct interactions with their audiences.

SIGMIS-CPR '19, June 20–22, 2019, Nashville, TN

ACM ISBN 978-1-4503-6088-3/19/06…$15.00
https://doi.org/10.1145/3322385.3322388

We are the first to begin to discover and enumerate motivating factors for cybersecurity professionals who serve as advocates. By understanding these motives, we begin to form a picture of how to attract and retain those who would be successful in the advocate role. To that end, we discuss implications for framing cybersecurity advocacy as a professional role fueled by these motivations and how these incentives may be maintained across advocates' careers.

2 RELATED WORK

We turn to past research on professional work motivation to provide context for our study. Work motivation is "a set of energetic forces that originates both within as well as beyond an individual's being, to initiate work-related behavior, and to determine its form, direction, intensity and duration" [23]. Motivation is often described in terms of being either intrinsic or extrinsic. Intrinsic motivators arise from an individual's feelings about a work activity and are inherent within the work itself [1]. These motivators can include interest, enjoyment, or feelings of accomplishment. Conversely, people are extrinsically motivated when they do the work in order to "obtain some goal that is apart from the work itself" [1]. Within the workplace, extrinsic motivators may include recognitions and monetary compensation.

While intrinsic and extrinsic motivators do interact, psychologists have found that intrinsic motivators most positively impact employee performance, creativity, and job retention [13]. In fact, offering excessive extrinsic rewards for work that is already intrinsically rewarding can be detrimental and lead to a decrease in overall motivation [13].

Work motivations of technology professionals were first explored within information technology (IT) and information systems (IS) fields. Based on Shein's career anchors (factors that give stability and direction to a person's career) [24], Crepeau et al. [9] identified significant IS worker anchors that included identity, service, and variety. Over 10 years later, Sumner and Yager [26] perhaps captured the evolving landscape of IT work, finding that the most compelling anchors for IT workers included organizational stability and variety, while identity, competence, creativity, and autonomy were viewed as less important.

Others explored motivation through lenses other than career anchors. Thatcher et al. [27] revealed that intrinsic motivators positively affected IT workers' job attitudes and suggested that further research is needed to identify nuances in motivation among different job types. Lounsbury et al. [20] found that disposition to teamwork and the motivation to achieve were positively related to both job and career satisfaction. Blum [5] explored gender-specific motivations for entering the computer science field, which is a primary feeder discipline into cybersecurity. He found that for men, computer science was seen as an interesting, fun discipline. Women, however, viewed it more as a means to achieve a socially motivated purpose.

Subsequent studies built upon this work to explore motivators within the much newer cybersecurity field. Chai and Kim [7] and Bashir et al. [4] identified security skill self-efficacy as a strong motivator for attraction to cybersecurity careers. Grounded in a literature survey, Dawson and Thomson [10] suggested that the future cybersecurity workforce should include those with a love of learning, a strong desire to work in teams, and sense of civic duty. Our previous paper on advocate characteristics revealed service orientation and an interest in incorporating diverse disciplines within the work [16].

Others focused on retention. Burrell et al.'s [6] analysis of focus groups of government cybersecurity employees implied that intrinsic motivators are more effective than extrinsic motivators for retention and success in the public sector since government institutions cannot compete with private-sector salaries. An industry survey [18] revealed that over 60% of cybersecurity job seekers desire to work in a job where they are empowered and can protect data and people, while roughly half were motivated by salary.

While this literature provides valuable insight, it is unclear as to whether these same motivations apply to cybersecurity advocates. While their jobs possess similarities to those of other cybersecurity and IT workers, advocates play a unique role. For example, like their counterparts they must possess technical expertise, but their main focus is not on technology administration or oversight. They must be skilled in the art of influence, but are not technical sales representatives or marketers. Our research discovers where advocates' motivations are similar and where they differ from those performing other cybersecurity roles and how these distinctions might influence advocate recruitment and retention techniques.

3 METHODOLOGY

We conducted semi-structured interviews of 28 cybersecurity professionals who performed advocacy tasks as a significant component of their jobs[1]. We followed a research approach inspired by Grounded Theory in which data collection and analysis are conducted concurrently, with analysis influencing decisions on future data collection [8].

The study was approved by our institutional review board with participants providing informed consent and receiving no compensation. To protect confidentiality, data associated with a participant were assigned a code (e.g., P16).

3.1 Recruitment and Participants

We initially recruited from researcher contacts and internet searches those who self-identified as security advocates. We then considered snowball recommendations that allowed interviewees to identify other advocates. Our definitional boundary of the cybersecurity advocate role continued to evolve and guided subsequent recruitment as interviews progressed. To ensure the representation of a broad range of advocacy contexts, we purposefully selected individuals who performed different types of advocacy (e.g., security awareness training, security consultation) who worked in a variety of sectors, and who served different types of audiences. This resulted in a collection of information-rich cases [22].

To guide recruitment, we practiced theoretical sampling throughout data collection [8]. We recruited participants four or five at a time. The subsequent group of potential participants was then selected to include those who might be able to provide additional or different insights on areas of interest surfacing from the analysis of the preceding set. For example, when several participants raised

[1] This section is summarized from the methodology sections of previous papers [15, 16]

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

gender diversity concerns, we subsequently recruited additional female participants.

Table 1 provides an overview of relevant participant demographics, with some roles generalized to preserve anonymity. Our study sample consisted of 10 female and 18 male professionals with all having at least five years of experience in the cybersecurity field, and 22 having more than 10 years. Fourteen participants had at least one non-technical degree (e.g., philosophy, communications, business, and law), with 11 of those having no formal technical degrees. Participants had worked within diverse sectors throughout their careers, including government, private industry, education, and non-profit organizations, with most having experience in more than one sector. Ten participants advocated security mainly external to their organization, three were focused within their organizations, and 15 advocated both internally and externally.

3.2 Data Collection

Interviews lasted on average 45 minutes and were audio recorded and transcribed. Interview questions addressed work practices, professional motivations, rewards and challenges, characteristics of successful advocates, and advocacy techniques. The interview protocol is included in Appendix A. All but one participant also completed an online demographic survey that included career background information.

The semi-structured nature of the interviews allowed for follow-on questions and the elicitation of rich data. The interview structure was ordered enough to facilitate cross-participant comparison, but open-ended enough to permit participants to raise themes we had not imagined in advance. Interviews were conducted until we reached theoretical saturation, the point at which no new ideas emerged from the data during our concurrent analysis [8]. Since the goal of qualitative research is rich, holistic contextual understanding, and not predictive generalization, the attainment of theoretical saturation indicated that we had reached an appropriate number of interviews [8]. This number also well exceeded the recommended minimum sample size of 12-20 for identifying themes in qualitative interviews [14].

3.3 Analysis

Methods for data coding and analysis were informed by Grounded Theory, which allows for an organic emergence of themes [8]. Each author initially reviewed five interviews and conducted inductive, open coding to label and discover meaning. We later met several times to discuss concepts identified from the interviews and begin to develop a codebook. The first author then used the codebook to recode the initial five interviews to align, and then deductively code the remaining interviews. As analysis progressed and additional concepts emerged, we made adjustments to the codebook. We then evolved our analysis into axial coding (the recognition of relationships among codes), captured emerging ideas within analytic memos, and identified core concepts (selective coding) [8].

4 FINDINGS

In this section we report on motivations for cybersecurity advocates as mentioned by study participants. These motivators contributed to a great passion for advocacy work, as expressed by 15 participants. One participant commented that advocacy *"became kind of calling over the years for me" (P04).* Another, from a non-profit, reflected on her work: *"I love it. It actually has a lot of the different gratifying qualities I enjoy in a job" (P24).*

No appreciable differences were observed among the various demographic groups (e.g., gender, formal education). In this section, we also provide participant counts to reflect frequency of concepts mentioned during the interviews. However, because our analysis was focused on identifying centrality of data codes to concepts, we caution the reader against making inferences beyond frequency.

4.1 Personal Interest

The belief that cybersecurity is a challenging and an *"intellectually exciting" (P16)* field was a motivator for 19 participants. A graphic designer-turned-advocate commented that the security awareness profession is *"like a giant puzzle. And challenges are my thing. I like being able to put effort into something that's creative and interesting and different" (P28).*

Since the cybersecurity field is relatively young and quite dynamic, it offers opportunities for innovation, which appealed to participants. One said, *"I'm attracted to areas where there's new things to do,... where it's not really established, where I get to solve a new problem, and solve a new problem that matters" (P19).* A security awareness program director at a public university commented,

> *"It's ever-evolving... I find it to be a challenge because threats and vulnerabilities in all environments... are always there and they're always becoming more sophisticated... That's what motivates me to stay in" (P14).*

We also identified an interest in interdisciplinary work (7 participants). The nature of advocates' work is multi-faceted in that it must consider challenges *"at the people level, at the business level, the strategic level" (P27).* An advocate with a strong cybersecurity background fell into security awareness when she became exposed to the human aspects of security: *"the power of human motivation fascinates me. And I think if I wasn't doing this, I'd probably be a behavioral psychologist or something" (P21).*

Another participant worked at the intersection of cybersecurity and usability:

> *"I tend to have an interest in things that are interdisciplinary and kind of at the border of different things... So, this combination of the technical and human factors interests me, and I guess I seem to be good at it, so that encourages me to want to do more" (P07).*

4.2 Sense of Duty

Almost all participants (26) exhibited an acute sense of duty and service. For example, a participant who works to influence public policy stated, *"I think we're making the world a better place" (P06).*

At the core of sense of duty was the perception of the importance of advocacy work. Although cybersecurity problems may seem overwhelming, participants thought that potential personal, economic, and national security consequences were too significant for them not to act. As one participant said, the role of cybersecurity is important *"in our future economy and in the management of social*

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

Table 1: Participant Demographics

ID	Gender	Role	Sector	Education	Audience
P01	M	Security analyst	***G***	T,N	B
P02	M	Professor	***E***,G,I	T,N	B
P03	F	Computer scientist	***G***,I	T	B
P04	M	Security evangelist	***N***,G	T	B
P05	M	Security researcher	***I***,G	T	B
P06	M	Director	***N***,G,E,I	N	B
P07	F	Senior technologist	***G***,***E***,I	T	E
P08	M	Security consultant	***I***	N	E
P09	M	Training director	***E***,G	N	E
P10	M	Instructor, consultant	***I***,***E***,G	T	E
P11	M	Director	***N***,I	N	E
P12	M	Security engineer	***I***,***E***,G	T	E
P13	M	Security engineer	***I***	U	I
P14	M	Security awareness director	***E***,G	N	B
P15	F	Director	***N***,***E***,I	N	B
P16	M	Computer scientist	***G***,E,I	T,N	I
P17	M	Researcher	***I***	T	E
P18	M	CIO	***E***	T	B
P19	F	Senior Architect	***I***	T	I
P20	M	Professor	***E***,***G***	T	E
P21	F	Company co-founder	***I***,G	T	E
P22	M	Security researcher	***I***, E	T	B
P23	F	Security consultant	***I***,E	N	B
P24	F	Director	***N***	N	E
P25	F	Deputy CIO	***G***,I	N	B
P26	F	CISO	***G***,I	T	B
P27	M	Director	***N***,I	N	B
P28	F	Security awareness director	***I***,E	N	B

Sector (***Current***,Past): E=Education, G=Government, I=Industry, N=Non-profit; **Education**: T=Technical degree, N=Non-technical degree, U=unknown/not reported; **Audience**: I=Internal to own organization, E=External to own organization, B=Both internal and external

issues like privacy and the way that we interact as social creatures across society. Cybersecurity is central to all of that" (P04).

Several advocates saw the potential of poor cybersecurity resulting in the loss of lives. One participant warned, *"If we don't get computers right, people are going to starve. And right now, we're not doing a good job" (P16).* Another advocate who leads a non-profit discussed his group's motivation:

> *"We had said we want to save lives through security research. It was really wherever bits and bytes meet flesh. That could be cars, medical devices, industrial control systems. But everyone's so focused on data and the confidentiality of data… We're spending nothing on our life and limb" (P11).*

Precipitated by the importance of the work, participants had a keen desire to help people navigate the dangers and complexities of the cyber world. A usable security champion commented, *"I like making people's lives easier" (P03).* Another participant said *"There's huge difference between my job satisfaction level when I know what I'm doing, day-in and day-out, is out there helping ultimately the citizens and the general population" (P26).*

All participants attempted to address the gap in security knowledge by playing formal or informal educator roles. Fifteen served as educators/mentors to future and current security professionals. A veteran advocate stated:

> *"I'm really conscious of my role as an old guy in this, a pioneer, someone who's got a lot of history. And so there's an excitement to that, to feel that you've seen a lot of things happen, made a lot of mistakes you get to convey to other people… So, feeling responsible. I feel the role there. It feels like I'm supposed to be doing this" (P04).*

Other advocates educated less-technical audiences. For example, one participant extended her advocacy responsibilities outside of work:

> *"I actually spend a significant time of my personal life … educating teachers and working with the old lady gang on my block to get them to understand security so that they're not in a position where they have to deal with some criminal stealing their information or stealing their hard-earned money" (P23).*

Another discussed his upcoming talk to a local community group:

> *"I'm trying to tune the message. What should citizens care about?… They're all great people, but they're not going learn what I've learned. So what is it that I can tell them that will*

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

help them, to get their attention, to cause them to change behavior?... there's a lot of great technologists in this business, but based on technology, we're not going to change people's behavior – well, only in niches. So, how do we put our work in other people's context" (P04)

4.3 Self-Efficacy

Self-efficacy, a belief in one's own ability to accomplish a task or exert control in specific situations [3], can be an important motivator [4, 7]. A large part of self-efficacy is self-confidence that one has the necessary knowledge and skillsets.

All participants exhibited self-confidence in their abilities to effectively perform advocacy tasks. This confidence was gained through years of experience and continuous learning. An advocate reflected on how his involvement with operational and threat intelligence organizations contributed to his effectiveness: *"Those two perspectives help me bring some unique value to the problem" (P05).*

Other advocates expressed confidence in their non-technical "soft" skills which were often deemed as more important than technical skills in advocacy work. For example, the ability to translate technical topics into layman's terms was noted as an important competency by 24 participants. When giving presentations to non-technical audiences, one participant commented, *"I seem to have the magic power to make these things make sense" (P08).*

4.4 Evidence of Impact

Since the goal of cybersecurity advocacy is behavior change, evidence that an advocate's recommendations have been understood, concurred with, and acted upon served as strong motivators for our participants and contributed to self-efficacy (23 participants). One participant said the most rewarding part of her job was *"seeing the impact, seeing some difference was made ... however minor or modest" (P19).*

We must also note that, although most advocates focused on successes, others were more forthcoming about challenges. When asked about his professional motivation, one participant commented on his frustrations:

"I'd love to say that it's to help people fight the good fight and make the world a better place. And it is certainly part of that, but I have to say, it isn't all that because, if it was, I would be very discouraged... We as an industry [are making the same mistakes] we've been making forever" (P10).

In the following subsections, we describe ways in which advocates realized the impact of their work as categorized by the sources providing evidence of impact.

4.4.1 Organizational Impact. Impacting organizations (mentioned by 10 participants) can be especially challenging since organizational barriers to cybersecurity, such as security culture, can be difficult to change. Therefore, a positive shift in attitude often provided hope of future behavior change. One participant commented:

"The impact isn't always... an easy thing to quantify in this field. So, I think a lot of times it's when the organization starts showing the passion, starts showing and are being responsive to the ideas you're trying to suggest. That can be very rewarding" (P05).

Another advocate, who had spent a substantial amount of his career conducting vulnerability assessments, talked about a feeling of accomplishment coming with

"the knowledge that the people were really onboard and believed what it was they were doing, believed that what we were telling them was important, and had the right guidance and the right authority to be able to move forward with it" (P01).

One participant spoke about how the effectiveness of a security awareness program might be measured by incremental shifts in security culture:

"It goes beyond just behavior, but more on the culture. So when you're talking to people, if they have a positive attitude about cybersecurity, a positive attitude about the cybersecurity team, if they feel like their behaviors have a positive impact, that's the first real big indicator that you've got a long-term win. Now the problem you have is changing culture's a three to ten year process" (P09).

4.4.2 Individual Impact. Eleven participants talked about the satisfaction of educating and making an impact on individuals. One advocate said:

"I always get really excited when I can just tell people have learned something. So when I see that little lightbulb come on... when I get confirmation that something I've said makes a difference, I get really excited" (P23).

A participant noted that, after giving a presentation,

"I definitely like interacting with people and people telling me, 'Wow, I learned something. I can use this. I'm going to change what I do, and this will help me.' I find that rewarding" (P07).

Another commented that he feels energized

"when I'm working with an adult, and they have the 'Eureka!' moment. They're struggling with understanding something, and you kind of sit next to them, and then they get it and you can see it in their eyes. Often a high-five moment happens" (P10).

4.4.3 Policy Influence. Five participants were able to influence broad-reaching cybersecurity policies. One participant who had worked in the government sector talked about how his organization had *"shaped the spending of hundreds of millions of dollars and the behavior of thousands" (P04).* P06's non-profit successfully lobbied for a substantial paradigm shift in cybersecurity public policy within the United States. P11 influenced medical device policy that prevented serious vulnerabilities from claiming lives.

4.4.4 Transfer of Knowledge. A deeply satisfying aspect of the job is the observation of the target audience taking the information they had learned and transferring that to others, as was mentioned by eight participants. A participant told a story about how an elderly neighbor whom she had taught cybersecurity best practices taught her son a lesson:

"Her son was buying a house, and she kept telling him, 'Don't email that stuff. Don't email your personal stuff.' Well, probably

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

about two weeks into purchasing the home, he realized that his email had been completely compromised…She saved him probably three million dollars because what happened is, in the middle of setting up escrow, someone asked for a bank account number. It wasn't anyone from the real estate firm" (P23).

Three participants discussed their roles in helping good security habits blur across the home/work divide. For example, one participant remarked,

"if you can train them with their home life and help them there, too, they can hopefully bring those behaviors to work and bring that sense of awareness up. So a lot of the organizations share those personal, consumer-focused resources with their employees so they can keep their families safe at home, too" (P24).

A security awareness program director provided another example:

"We had a videotape…One of the presenters…had been kidnapped. She had been social engineered by a man online…And I know some of the people took their laptops home, logged in, and made their kids watch that. So I think that's great, too, when the information that you're giving at work, people are sharing with friends and family, too" (P28).

4.4.5 Metrics. Nine participants viewed metrics as motivating evidence that they were on the right path with their approaches. Metrics mentioned by participants included how many people accessed publications and videos, the number of newsletter subscriptions, attendance of security events, the growth of non-profit membership, and statistics showing improvement in security behaviors. A participant discussed the importance of metrics in showing success of an organization's security awareness program: *"initially you may have to measure success by some specific behaviors such as phishing, exposing of sensitive data, use of ID badges and things like that" (P09).*

Advocates monitored these metrics as one indicator of both the reach of their message and effectiveness of a communication channel. For example, one participant noted that one of her talks had been recorded and posted online and had *"been viewed like two million times…There's a lot of bang for that buck" (P07).* A participant whose organization produced cybersecurity implementation resources said:

"We create specific things, sort of products to give away, ideas and papers. So part of that is it comes with some natural mechanism now to calibrate feedback. How often is it downloaded, how often is it referenced?…So we see lots of interest worldwide. We can count how many tens of thousands of downloads there have been" (P04).

4.4.6 Praise. Praise is an extrinsic motivator that, when sincere, can significantly contribute to intrinsic motivation. Fourteen participants felt valued after receiving positive feedback from their audience. Some feedback was more formal, often obtained through surveys. For example, a non-profit advocate commented:

"We have teams to reach out to adopters of our work, talk to them, ask them questions through surveys, write down use cases if they're willing, that sort of thing …feedback is not an issue. We get a lot of that. I'd say it's overwhelming positive" (P04).

Informal feedback, such as face-to-face comments and email, seemed to be most personally satisfying. A participant who advocates to lawyers talked about one member of his audience saying, *" 'This is amazing! So many lawyers don't understand this.'…it's wonderful to get those sorts of reactions" (P08).* A security awareness director at a university remarked:

"I'll get stuff directly if I run an event of some kind, an awareness event, whether it's a conference or just a small brown bag session. I will receive emails from attendees saying, 'Hey, this was extremely helpful. I was unaware of XYZ component of data privacy. Or HIPAA [Health Insurance Portability and Accountability Act] privacy.' Whatever the topic might be. So the feedback that I receive in many instances is informal feedback" (P14).

4.5 Comradery

When asked about the rewards of their jobs, seven participants discussed their enjoyment working with others in the field. Whereas this motivation is not unique from other professions, it highlights the importance of a sense of belonging and collaboration in the cybersecurity advocate role and runs counter to commonly-held stereotypes of cybersecurity being a solitary profession [25].

A security course instructor and consultant commented, *"most of my very close friends are in this industry. I love to spend time with them thinking good thoughts" (P10).* A security evangelist at a non-profit described the benefits of working with a large group of volunteers to produce security guidance:

"It's a business full of really bright people, lot of diverse, creative, smart people of good will…So I love that part of it, the sort of community, this collaboration…it's something that's personally satisfying" (P04).

The interactions advocates have with others in the field are not just personally satisfying but are a necessary component of the job. Because of the distinctly dynamic nature of the cybersecurity field in which major developments can occur on a daily basis, advocates rely on a symbiotic relationship of receiving and contributing information within their personal networks. For example, an advocate who is active in a security awareness community commented, *"not only am I always giving to the community, I'm listening to the community…So, I always understand the latest and greatest risk from a human side" (P09).*

4.6 Awards and Monetary Compensation

Only five participants mentioned official recognitions or monetary compensation as motivators. One participant, referencing her team's best website award, said, *"I love it when my team is recognized" (P26).* Another was motivated by continued research funding: *"if they didn't like what you were doing, and they didn't think there was value, they wouldn't continue to fund you. And so, our funding's pretty stable" (P03).*

When asked about the rewards of his work, one advocate first mentioned the fun he has educating youth about cybersecurity and

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

the gratitude of his clients. However, he was then quite frank when he also included financial reward:

> *"As long as you're not cheating people or doing something dishonest, if you're providing real value, the way people indicate that you're providing real value to them is by paying you. So the more that they pay you, the more value you're providing... So I know that might sound crude, and maybe I should be more noble,... but I always realize at the end of the day, I gotta make payroll" (P10).*

5 IMPLICATIONS

Our findings confirm many of the same motivations as IT and cybersecurity professionals identified in the prior literature, such as sense of duty/service, self-efficacy, working with others, and interest, although sometimes to different degrees. We also similarly found an emphasis on intrinsic motivations (e.g., personal interest and sense of duty) inherent in the work of cybersecurity advocacy and its immediate goal of enacting positive behavior change. However, even extrinsic motivators (e.g., praise and compensation) were contributors to intrinsic motivation. For example, we observed that most participants were not ego-driven, so praise was viewed in the context of reinforcing self-efficacy and sense of duty.

Despite the similarities, we also recognize the uniqueness of the cybersecurity advocate role versus traditional IT or cybersecurity professionals, such as analysts, administrators, and system architects. To be effective, advocates must be more socially-oriented and skilled in persuasion and communications then their non-advocate colleagues [16]. They appear to be driven by a sense of duty and evidence of real impact, and often have larger spheres of influence. We therefore see the need for a more nuanced approach to feature advocate motivators in recruitment and retention by 1) advertising cybersecurity as a profession that has the potential to fuel these motivations via an advocacy role, 2) increasing motivations by providing opportunities for traditional cybersecurity professionals to progress into advocate positions, and 3) sustaining motivation for current advocates by documenting impact and providing energizing feedback.

5.1 Recruitment

When marketing cybersecurity positions having advocacy duties, in addition to touting the work as interesting and challenging {supported by section 4.1}, there should be emphasis on the important service to individuals and society {4.2}. For example, when recruiting advocates for jobs in the public sector, salaries can seldom compete with those in private industry, so appealing to motivators like a sense of civic duty and national pride may be especially helpful in attracting qualified individuals [10]. Service orientation of the work may also appeal to currently underrepresented populations in the cybersecurity workforce who may perceive cybersecurity as having no social benefit [25] or women who desire a career with a socially motivated purpose [5].

The opportunity to work collaboratively with talented, diverse people from multiple disciplines should also be highlighted {4.1}. This emphasis may counter a lack of awareness of the breadth of opportunities available in security careers [12] and belief that only those with deep technical skills can be successful [11, 12]. The interdisciplinary framing might help attract individuals from other fields who possess important non-technical skills and unique perspectives and encourage a greater sense of self-efficacy {4.3}. Additionally, an emphasis on the value of diversity may encourage participation of women and minorities who otherwise may be deterred by the stereotype of a white male, hacker-dominated workforce [2, 11].

5.2 Retention

Due to the dynamic nature of cybersecurity, organizational challenges, and human nature, the work of advocates can sometimes be daunting and thankless, requiring perseverance and resilience. In addition, individuals qualified for cybersecurity advocate positions possess a valuable blend of skills [16, 19], so are in danger of being recruited away by others. Therefore, special emphasis should be placed on their retention.

To aid in retention, foster advocate motivation, and encourage progression of current professionals into advocate roles, we propose the following recommendations for employers.

(1) Learn to recognize those who are doing advocacy work within the organization, even if in the background. Offer sincere praise and feedback about their successes (even if minor) and tout their mix of technical and non-technical skill. Provide opportunities to assume more responsibility for security promotion activities. {4.3, 4.4}
(2) Provide ample opportunity for advocates to receive direct feedback from their audience (face-to-face especially) about their efforts. Implement mechanisms to measure effectiveness and value of advocacy approaches. {4.2, 4.4}
(3) Support advocates in trying innovative approaches. {4.1, 4.3}
(4) Encourage advocates to participate in collaborative and information sharing opportunities with others working in related areas. {4.1, 4.5}
(5) Clearly communicate to the workforce that advocates are supported by leadership as important contributors in protecting people, systems, and information. {4.2, 4.3}
(6) Arm advocates with professional development and continuous learning opportunities that can aid them in their jobs. This learning should address the interdisciplinary nature of cybersecurity and include organizational, social, and technical aspects of cybersecurity. {4.1}
(7) Be cautious with offering excessive extrinsic incentives as these may interfere with intrinsic motivation. However, try to promote and pay advocates commensurate with the value they bring to the organization. If that is not possible, provide advocates with clear feedback about the importance and value of their work. {4.2, 4.4, 4.6}

6 LIMITATIONS

Our study is limited in that interviews are commonly subject to self-report and social desirability bias in which participants may adjust their answers to appear more acceptable to the interviewer. These biases may have particularly been a factor when so few participants mentioned monetary compensation as a motivator. In addition, although an interview protocol was used to ensure consistency across several predetermined topics, the semi-structured

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

interviews allowed for exploration of ideas that may not have been discussed equally in all interviews. Bias and consistency concerns were primarily mitigated by the diversity of our sample and constant comparison method of our analysis.

As the first to look at cybersecurity advocacy, we are discovering variables of interest that can be validated in future studies. For example, follow up efforts are underway to observe cybersecurity advocacy in practice within organizations.

7 CONCLUSION

Cybersecurity advocates serve as important enablers to security adoption. Our study is the first purposeful effort to learn about their work motivations. Most critically, we suggest ways to leverage these motivations in cybersecurity advocate recruitment and retention efforts to better position the cybersecurity workforce to meet the challenges of the future.

DISCLAIMER

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their comments that helped improve the quality of this paper.

REFERENCES

[1] Teresa M. Amabile. 1993. Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review* 3, 3 (1993), 185–201.
[2] Sharmistha Bagchi-Sen, H. Raghav Rao, Shambhu J. Upadhyaya, and Sangmi Chai. 2010. Women in cybersecurity: A study of career advancement. *IT Professional* 1 (Jan. 2010), 24–31.
[3] Albert Bandura. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84, 2 (1977).
[4] Masooda Bashir, Colin Wee, Nasir Memon, and Boyi Guo. 2017. Profiling cybersecurity competition participants: Self-efficacy, decision-making and interests predict effectiveness of competitions as a recruitment tool. *Computers & Security* 65 (2017), 153–165.
[5] Lenore Blum. 2001. Women in computer science: the Carnegie Mellon experience. *women@ scs2* 2, 1 (2001).
[6] Darrell Norman Burrell, Maurice Dawson, William Quisenberry, Aikyna Finch, and Angelique Goliday. 2012. An exploration of government talent management strategies for information assurance and cyber-security employees in organizations with public health and safety missions. *Journal of Knowledge & Human Resource Management* 4, 8 (2012).
[7] Sang-Mi Chai and Min-Kyun Kim. 2012. A Road To Retain Cybersecurity Professionals: An Examination of Career Decisions Among Cybersecurity Scholars. *Journal of The Korea Institute of Information Security and Cryptology* 22, 2 (2012), 295–316.
[8] Juliet Corbin and Anselm L. Strauss. 2015. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (4th ed.). Sage, Thousand Oaks, CA.
[9] Raymond G. Crepeau, Connie W. Crook, Martin D. Goslar, and Mark E. McMurtrey. 1992. Career anchors of information systems personnel. *Journal of Management Information Systems* 9, 2 (1992), 145–160.
[10] Jessica Dawson and Robert Thomson. 2018. The Future Cybersecurity Workforce: Going Beyond Technical Skills for Successful Cyber Performance. *Frontiers in Psychology* 9 (2018).
[11] Katharine D'Hondt. 2016. *Women in Cybersecurity.* Master's thesis. Harvard University, Cambridge, MA.
[12] Matthew D. Gonzalez. 2015. Building a Cybersecurity Pipeline to Attract, Train, and Retain Women. *Business Journal for Entrepreneurs* 3 (Sep. 2015).
[13] Richard A. Griggs. 2010. *Psychology: A concise introduction.* Macmillan.
[14] G. Guest, A. Bunce, and L. Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18, 1 (2006), 59–82.
[15] Julie M. Haney and Wayne G. Lutters. 2018. "It's Scary…It's Confusing…It's Dull": How cybersecurity advocates overcome negative perceptions of security. In *Proceedings of the Symposium on Usable Privacy and Security.* USENIX, Baltimore, MD, 411–425.
[16] Julie M. Haney and Wayne G. Lutters. 2018. Promoting skill and discipline diversity in cybersecurity advocacy. https://cdn.website-editor.net/22097006d5ba4ddbb1a13216c1bd98ca/files/uploaded/SP-JoC-18-05002.pdf. *Emerging Trends in Cybersecurity: Journal of Cybersecurity - Online* (October 2018).
[17] ISC2. 2018. Cybersecurity professionals focus on developing new skills as workforce gap widens - ISC2 Cybersecurity Workforce Study. https://www.isc2.org/-/media/ISC2/Research/2018-ISC2-Cybersecurity-Workforce-Study.ashx.
[18] ISC2. 2018. Hiring and Retaining Top Cybersecurity Talent. https://www.isc2.org/-/media/Files/Research/ISC2-Hiring-and-Retaining-Top-Cybersecurity-Talent.ashx.
[19] Ray Lapena. 2017. Survey Says: Soft Skills Highly Valued by Security Team. https://www.tripwire.com/state-of-security/featured/survey-says-soft-skills-highly-valued-security-team/.
[20] John W. Lounsbury, Lauren Moffitt, Lucy W. Gibson, Adam W. Drost, and Mark Stevens. 2007. An investigation of personality traits in relation to job and career satisfaction of information technology professionals. *Journal of Information Technology* 22, 2 (2007), 174–183.
[21] National Initiative for Cybersecurity Careers and Studies. 2019. Glossary. https://niccs.us-cert.gov/about-niccs/glossary.
[22] Michael Q. Patton. 2015. *Qualitative research and evaluation methods* (4th ed.). Sage, Thousand Oaks, CA.
[23] Craig C. Pinder. 2014. *Work motivation in organizational behavior* (2nd ed.). Psychology Press.
[24] Edgar H. Schein. 1978. *Career Dynamics: Matching Individual and Organizational Needs.* Addison-Wesley, Reading, MA.
[25] Rose Shumba, Kirsten Ferguson-Boucher, Elizabeth Sweedyk, Carol Taylor, Guy Franklin, Claude Turner, Corrine Sande, Gbemi Acholonu, Rebecca Bace, and Laura Hall. 2013. Cybersecurity, women and minorities: Findings and recommendations from a preliminary investigation. In *Proceedings of the ITiCSE Conference on Innovation and Technology in Computer Science Education.* ACM, Canterbury, UK, 1–14.
[26] Mary Sumner and Susan Yager. 2004. Career orientation of IT personnel. In *Proceedings of the 2004 SIGMIS Conference on Computer Personnel Research.* ACM, Tucson, AZ, USA, 92–96.
[27] Jason Bennett Thatcher, Yongmei Liu, and Lee P. Stepina. 2002. The role of the work itself: An empirical examination of intrinsic motivation's influence on IT workers attitudes and intentions. In *Proceedings of the 2002 SIGMIS Conference on Computer Personnel Research.* ACM, Kristiansand, Norway, 25–33.
[28] Verizon. 2018. 2018 Data Breach Investigations Report. https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_execsummary_en_xg.pdf.

A INTERVIEW PROTOCOL

Items in **bold** are those most relevant to the theme of motivations discussed in this paper.

(1) Can you tell me about what you do in your job?
(2) How did you come to do this type of work?
(3) What motivates you to do this work?
(4) What do you think is the importance of your role in promoting security?
(5) How is your work is valued by others?
  (a) What kind of feedback do you get?
  (b) Can you talk about any times when you felt that your work wasn't appreciated?
(6) What do you think are qualities or characteristics of people who are successful in promoting security?
(7) Have you had experiences with or know of security advocates who you don't think were particularly effective? What

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

was it about them or what did they did or did not do that contributed to their ineffectiveness?

(8) Through what means do you promote security? For example, conferences, invited talks, blogs, social media, articles, client visits, face-to-face meetings, phone, email.
   (a) Which of those means do you think are the most effective? Why?

(9) What are your thoughts about whether or not you are reaching the right population of people and organizations?
   (a) What is preventing you from reaching the right people?
   (b) What do you wish you could do to reach the right population?

(10) How do you keep up with the latest in security?

(11) What do you find most rewarding about your work?

(12) What do you find most challenging or frustrating, if anything, about your role as a security advocate?

(13) What do you think are the biggest obstacles individuals and organizations face with respect to implementing security measures and technologies?

(14) What do you see as your role in helping organizations overcome these obstacles?

(15) Is there anything else you'd like to add with respect to what we've talked about today?

Haney, Julie; Lutters, Wayne.
"Motivating Cybersecurity Advocates: Implications for Recruitment and Retention."
Paper presented at SIGMIS-CPR '19: 2019 Computers and People Research Conference, Nashville, TN, United States. June 20, 2019 - June 22, 2019.

Proc. "Frontiers of Characterization and Metrology for Nanoelectronics: 2019" April 2-4, 2019, Monterey, CA
PP 85 to 87

Nondestructive and Economical Dimensional Metrology of Deep Structures

Ravi Kiran Attota [1,4], Hyeonggon Kang [2], and Benjamin Bunday[3]

[1] *Microsystems and Nanotechnology Division, NIST, Gaithersburg, MD 20899, USA*
[2] *Coppin State University, Baltimore, MD 21216, USA*
[3] *No affiliation, USA*
[4] *Corresponding author: Ravikiran.attota@nist.gov*

INTRODUCTION

Low-cost, high-throughput and nondestructive metrology of truly three-dimensional (3-D) targets for process control/monitoring is a critically needed enabling technology for high-volume manufacturing (HVM) of nano/micro technologies in multiple areas [1,2]. A survey of the typically used metrology tools indicates the lack of a tool that truly satisfies the HVM metrology needs of 3-D targets, such as high aspect ratio (HAR) targets and through silicon vias (TSVs). TSVs are a key component to enable 3-D stacked integrated circuits (3DS-IC), which themselves are key to extended scaling of integrated circuits and enabling heterogeneous integration [1]. Hence it is crucial to find suitable metrology solutions for truly 3-D targets such as HAR targets.

To be used in high volume manufacturing, a metrology tool must – in addition to providing statistically significant results [2] – be fast (high-throughput), low-cost, inline capable, automated, robust, easy to use, non-contact and non-destructive. The requirements for satisfactory measurement sensitivity and resolution have been identified in the International Technology Roadmap for Semiconductors (ITRS) and International Roadmap for Devices and Systems (IRDS) 2017 Edition: Metrology. All currently available tools have certain advantages and disadvantages. It is difficult to find a metrology tool that satisfies all the above-mentioned requirements, especially for metrology of 3-D/HAR targets. Here we show that through-focus scanning optical microscopy (TSOM) [3,4] could supplement currently available metrology tools in filling this gap for 3-D shape metrology. However, there is still room to improve optical tools [5].

THROUGH-FOCUS SCANNING OPTICAL MICROSCOPY (TSOM)

TSOM is a method that collects and preserves the entire through-focus optical intensity information in 3-D space using a conventional optical microscope. Developments in image acquisition techniques have significantly reduced the acquisition time for a set of through focus images to be as fast as a single conventional microscope image, making TSOM suitable for HVM [6]. A vertical cross-section extracted from this 3D data results in a TSOM image. D-TSOM images are generated by taking a pixel-by-pixel difference between two TSOM images obtained using two different targets. D-TSOM images expose small (down to sub-nanometer) differences hidden in nominally identical targets. The color patterns of D-TSOM images are usually distinct for different types of parameter changes and serve as a "fingerprint" for different types of parameter variations, but remain qualitatively similar for different magnitude changes in the same parameter. However, the magnitude of the optical content of D-TSOM images is proportional to the magnitude of the dimensional differences. The Optical Intensity Range (OIR, the difference between the maximum and the minimum optical intensity, multiplied by 100), provides a quantitative estimate of the difference between two images. The utility of D-TSOM is that the color pattern of the D-TSOM image is an indicator of the difference in 3D shape, while the magnitude of the OIR scales with the dimensional difference between the two targets.

Attota, Ravikiran; Kang, Hyeonggon; Bunday, Benjamin.
"Nondestructive and Economical Dimensional Metrology of Deep Structures."
Paper presented at Frontiers of Characterization and Metrology for Nanoelectronics (FCMN), Monterey, CA, United States. April 2, 2019 - April 4, 2019.

Proc. "Frontiers of Characterization and Metrology for Nanoelectronics: 2019" April 2-4, 2019, Monterey, CA
PP 85 to 87

DEMONSTRATION OF TSOM FOR PROCESS MONITORING

Application for High Aspect Ratio (HAR) Trenches

As a demonstration of the application of TSOM to HAR structures, trench targets in SiO_2 layer on a 300 mm Si substrate with nominally 100 nm CD and 1100 nm depth were studied [7]. A mosaic of D-TSOM images obtained (with central die as a reference target) for the entire wafer is presented in Fig. 1(a). From this, four major types of D-TSOM image patterns (Fig. 1(b)) can be identified with their corresponding FIB cross-sectional profile differences as shown in Fig. 1(c). Considering the information available in Figs. 1(b) and (c) as a library, a simple process monitoring procedure can be proposed as shown in Fig. 2, based on the following selected rules: If the OIR of the D-TSOM image is more than 12, reject the target, as the dimensional differences are more than the tolerable limits. On the lower side, if the OIR of the D-TSOM image is less than 7, accept the target as the dimensional differences are within the acceptable level. If the OIR value is in between 7 and 12, accept the production target if the profiles are symmetric (T1 and T3) and reject if the profiles are asymmetric (T2 and T4).

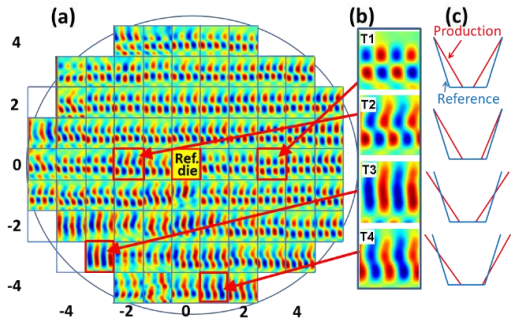


FIGURE 1. (a) A mosaic of the D-TSOM images obtained by subtracting the TSOM image of the central reference target from the TSOM images of the targets in the other dies (with color scale bar set to automatic). (b) Four major types (T1, T2, T3 and T4) of D-TSOM image color patterns are identified. (c) Schematic cross-sectional profile differences corresponding to (b) obtained from FIB cross-sectional analysis. Nominal pitch = 1,000 nm. Illumination wavelength = 520 nm, numerical apertures (NA) = 0.75, illumination NA (INA) = 0.25.

Row no. (Die location)	Production target (unknown Profile)	Library T1	Library T2	Library T3	Library T4	Possible D-TSOM image	Inferred shape difference type	OIR	Accept(✓) or Reject(⊗) status
		Correlation coefficient/match							
1 (-2,2)		**0.96**	0.69	0.66	0.61		T1	8.4	✓
2 (1,-4)		0.64	0.79	0.78	**1.00**		T4	8.0	⊗
3 (-2,1)		0.84	**0.85**	0.59	0.54		T2	6.3	✓
4 (4,2)		**0.90**	0.77	0.65	0.52		T1	13.6	⊗

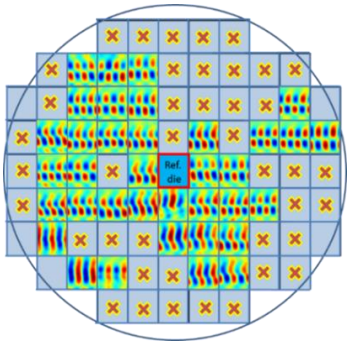


FIGURE 2. (Left) Proposed TSOM-based automated 3-D-shape process control method. The selected test production targets in column 1 are considered to have unknown 3-D-shape profile (column 2). Comparing the correlation coefficients of the D-TSOM images of the test targets with the library provides the best match (green boxes) from which the possible 3-D shape difference type can be inferred (column 8). Based on the type of 3-D shape difference and the magnitude of the dimensional difference (OIR), the process control decision of accept/reject status can be made. (Right) Dies marked by an X are rejected after applying the automated TSOM-based process control criteria. The rest of the dies are deemed acceptable.

Attota, Ravikiran; Kang, Hyeonggon; Bunday, Benjamin.
"Nondestructive and Economical Dimensional Metrology of Deep Structures."
Paper presented at Frontiers of Characterization and Metrology for Nanoelectronics (FCMN), Monterey, CA, United States. April 2, 2019 - April 4, 2019.

Proc. "Frontiers of Characterization and Metrology for Nanoelectronics: 2019" April 2-4, 2019, Monterey, CA
PP 85 to 87

Application for Through Silicon Vias (TSVs)

The premise of the TSOM approach is that the use of conventional optical microscopy components, as opposed to infrared reflectometry, may provide a cost-effective method for TSV metrology, especially for process control. In the example we present here, dimensional differences between three TSVs from three dies are evaluated by comparing the TSVs, pairwise, using D-TSOM images. The comparison can be evaluated numerically (OIR) and can also be displayed graphically. Figure 3 shows experimental D-TSOM data for TSVs with a nominal diameter and depth of 1 μm and 20 μm, respectively. The bar graphics above the image plots indicate which two TSVs are being compared in each plot; the values of the OIR indicate the overall magnitude of the dimensional differences between the TSVs being compared; and the colors indicate the spatial distribution of dimensional differences between each pair of TSVs. The similarity of the two differential images on the right indicates a similar pattern of dimensional differences between the pairs of TSVs compared. The OIR value of the image at the right is the highest, indicating the largest magnitude dimensional difference, while the lowest OIR value of the left image indicates the smallest difference. If a library of differential TSOM images corresponding to known dimensional differences were available, the actual dimensional differences corresponding to these image patterns could be evaluated.

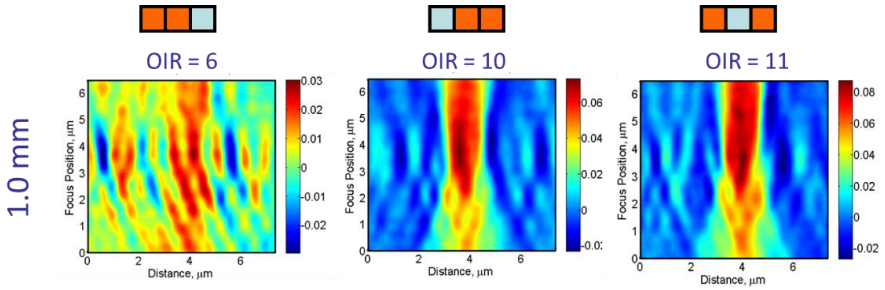


FIGURE 3. D-TSOM images representing the pairwise differences between TSOM images of three TSVs in three different dies.

Here we have briefly presented potential applications of TSOM for dimensional metrology of deep structures such as HAR targets and TSVs using a reference library. TSOM is a nondestructive, low-cost, and high-throughput 3-D shape process monitoring method that uses a conventional optical microscope. TSOM could fill a gap by satisfying the HVM metrology needs of not only HAR targets and TSVs, but also other types of truly 3-D targets for which 3-D shape process control/monitoring is needed, complementing other widely used metrology tools.

Acknowledgements: We thank Keana Scott, Richard Allen, Andras Vladar, John Kramar, and Victor Vertanian for their direct or indirect help.

REFERENCES

[1] V. Vartanian, R. A. Allen, L. Smith, K. Hummler, S. Olson, and B. Sapp, *J Micro-Nanolith Mem* **13** (1) (2014).
[2] B. Bunday, *Proc. SPIE* **9778** (2016).
[3] R. Attota and R. G. Dixson, *Appl Phys Lett* **105** (2014).
[4] R. Attota, R. G. Dixson, J. A. Kramar, J. E. Potzick, A. E. Vladar, B. Bunday, E. Novak, and A. Rudack, *Proc. SPIE* **7971**, 79710T (2011).
[5] E. Agocs and R.K. Attota, *Sci Rep* **8**, 4782 (2018).
[6] R.Attota, *J Biomed Opt* **23** (7), 19100 (2018).
[7] R.K. Attota, H. Kang, K. Scott, R. Allen, A. E. Vladar, and B.in Bunday, *Measurement Science and Technology* **29**, 125007 (2018).

KEYWORDS

TSOM, nondestructive process control, three-dimensional metrology, through-focus scanning optical microscopy, nanometrology, packaging metrology, high-throughput semiconductor metrology

Switching variability factors in compliance-free metal oxide RRAM

D. Veksler, G. Bersuker, A. W. Bushmaker
MTD, The Aerospace Corporation
Los Angeles CA
dmitry.veksler@aero.org

P. R. Shrestha, K. P. Cheung, J. P. Campbell
NDCD, National Institute of Standards and Technology
Gaithersburg, MD

***Abstract*— Switching variability in polycrystalline compliance-free HfO_2-based 1R RRAM is evaluated employing ultra-fast low voltage pulse approach. Changes in filament conductivity are linked to the variations of energy consumed in a switching process. This study indicates that variability is reduced (suppressed) in more resistive filaments.**

***Index Terms*-- Neuromorphic computing, RRAM.**

I. Introduction

Mobile neuromorphic computing (NC) systems introduce specific requirements to RRAM cells to be used as microelectronic "synapses". Sufficiently high stability and low variability of the memory states for the reliable gradual memory update is required in analog applications to implement adaptive synaptic changes. Formation and switching of the conductive filament in the metal oxides involves an atomic-level rearrangement that is a stochastic process resulting in high device-to-device and cycling variability [1]. Such stochasticity can be expected to reduce when redox processes are limited by a tighter control over the duration of the forming/switching operations and the oxide region where they take place. In this respect, the ultra-short pulse technique [2], which corresponds to actual circuitry operation frequencies, is an enabling tool to evaluate RRAM characteristics under use conditions. In this study, we focus on identifying operational factors, which can affect switching variability in hafnia-based filamental RRAM devices.

II. Measurements Setup

RRAM devices are formed by crossbar 50x50 nm MIM capacitors fabricated with an ALD polycrystalline 5nm HfO_2 film, overlaying oxygen scavenging Ti layer and TiN electrodes. Figure 1 show a setup, which delivers sufficiently short voltage pulses limiting energy dissipation and related redox processes that determines resistance of the formed filament and removes the need for a current compliance control.

The energy dissipated in the RRAM cell during forming can be calculated as:

$$E = \int_{t_0}^{t_0+t_{pulse}} V(t)I(t)dt \qquad (1)$$

This research was partially supported by The Aerospace Corporation Internal Research and Development Program.

where V(t) is the voltage pulse, I(t) is the current through the RRAM cell during the pulse. The distribution of the post-forming resistance values at a given pulse width (Fig. 2) is driven by the variation of these forming energies, Eq. (1), caused by device-to-device variability of initial (pre-forming) precursor conductive paths, rather than variability of the forming process itself. By tuning the pulse amplitude, state resistance can be gradually, via multiple pulses, changed to the desirable level, Fig. 3. Cyclic switching between two resistance states was performed by employing a feedback system allowing to stop and flip the polarity of the programming pulse when the desired conductance state of the memory cell is reached. While LRS shows tight distribution over switching cycles, the HRS fluctuates. However, the frequency and amplitudes of large HRS fluctuation remain limited (up to R_{max}) during continuous switching (Fig. 4). It implies that these fluctuations are not associated with permanent structural changes, and observed resistance variations reflect random reversible processes (their origin will be discussed elsewhere).

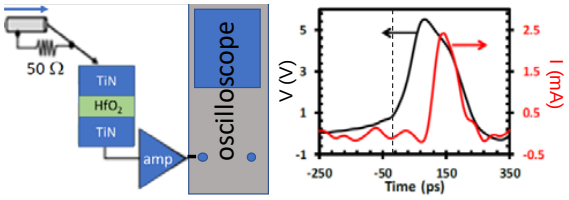


Figure 1. *Short pulse RRAM switching setup. Example of measured voltage and current signal during RRAM forming extracted from oscilloscope measurements.*

III. Variability and Energy Dissipation

To understand the variability drivers, we compare cycle-to-cycle conductance variations, along with the corresponding distributions of switching energies. Due to extremely low parasitic capacitance in this crossbar architecture, it was possible to directly measure switching current down to ns pulses and extract switching energies, Eq. (1). In the samples that drift, device D - Fig. 5a, and in those that are stable, device S - Fig. 5b, switching characteristics (nominally identical devices D and S were formed targeting different post-forming

Veksler, Dmitry; Bersuker, Gennadi; Bushmaker, A; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"Switching variability factors in compliance-free metal oxide RRAM."
Paper presented at 2019 International Reliability Physics Symposium, Monterey, CA, United States. March 31, 2019 - April 4, 2019.

resistance. The more conductive device D exhibits higher cycling variability of both filament conductance (Fig. 5) and switching energy (Fig. 6).

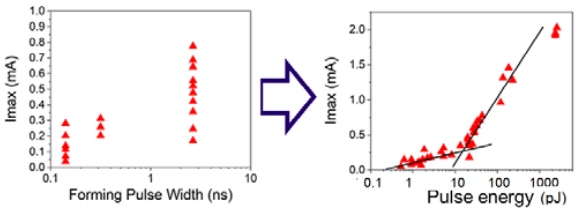


Figure 2. *(a) Device-to-device distribution of the post-forming I_{max} vs. pulse durations. (b) I_{max} vs. energy consumed during forming pulse, E_{Form}. Above a certain critical energy value, the filament conductance increases at a much higher rate as indicated by the line slopes.*

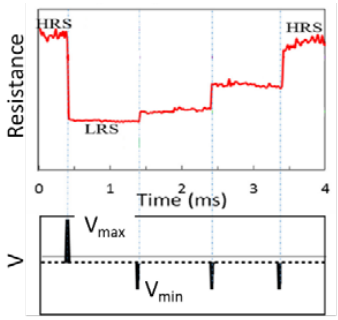


Figure 3. *Resistance switching can be performed in analog regimes using multiple pulses ($|V_{max}|$ and $|V_{min}| < 1V$)*

Since larger switching energy is consumed when the device is in a lower resistance state (higher current), variation in LRS conductance in the device D directly reflects on the variation of switching energy and, thus, they strongly correlate. The device with higher resistance, device S, demonstrates significantly tighter distributions of switching energies and conductance values in both HRS and LRS.

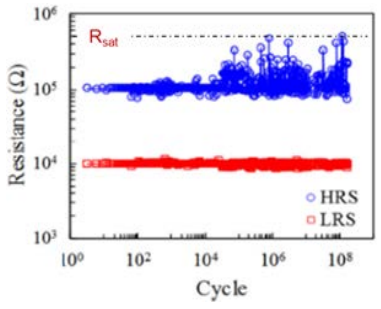


Figure 4. *Endurance: LRS and HRS resistances vs. a number of switching cycles (in logscale). HRS resistance fluctuations are bounded, maximum variation of resistance, R_{sat}, is observed approximately every 10000 cycles, fluctuations frequencies are constant throughout the entire switching cycle.*

Conductivity values in 10^3 SET/ReSET cycles of the devices D and S are plotted vs. SET/ReSET energies consumed in each of these cycles (calculated using Eq. 1) in Fig. 7. The same starting conductance states may lead to different conductivities after switching, as illustrated by an example of the ReSET switching data in Figs. 7,8. Higher starting LRS conductivities are associated with higher resulted HRS conductivities and switching energies (Fig. 8). It can be seen, however, that the variation of switching energy alone cannot account for the conductivity distribution at any given switching energy value. Indeed, the energy is determined by the contributions from both conductive states participating in interstate switching, when the voltage and duration of the switching pulses are hold constant, Eq. 1. In the ReSET process, LRS dominates: a current in this state determines the maximum consumed switching energies. In SET, a smaller portion of the overall switching time is associated with LRS, resulting in smaller SET energies in Fig. 7. Thus, the switching direction (SET or ReSET) also affects the outcome.

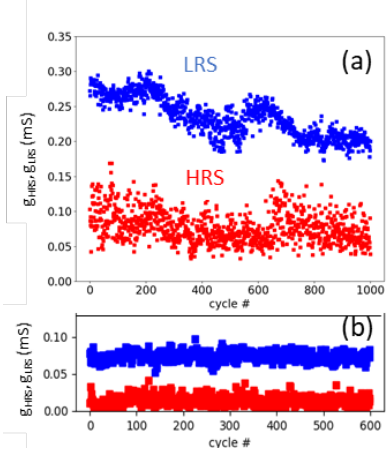


Figure 5. *LRS and HRS conductance in (a) device D with drifting switching characteristics and (b) device S with stable switching characteristics.*

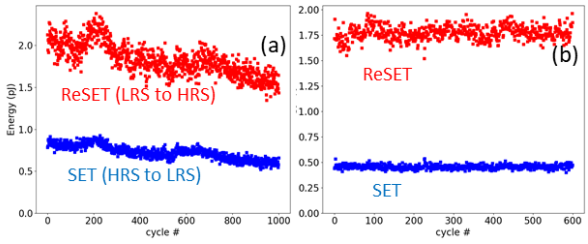


Figure 6. *Energy consumed in each SET and ReSET switching event in (a) device D and (b) device S.*

To identify the variability drivers, we start with the factors responsible for switching. Conductivity changes, that is differences between pre- and post- switching values: $\Delta g_{LRS} = g_{LRS} - g_{HRS}$ (SET) and $\Delta g_{HRS} = |g_{HRS} - g_{LRS}|$ (ReSET), are plotted vs. energies associated with the corresponding switching processes in Fig. 9. A much wider range of conductance

Veksler, Dmitry; Bersuker, Gennadi; Bushmaker, A; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"Switching variability factors in compliance-free metal oxide RRAM."
Paper presented at 2019 International Reliability Physics Symposium, Monterey, CA, United States. March 31, 2019 - April 4, 2019.

changes in device D can be linked not only to differences in switching energies, as expected in the case of varying resistance values (in Fig. 6), but also to the resistances of the starting states, from which the switching processes originate (large Δg spread-outs at each energy value).

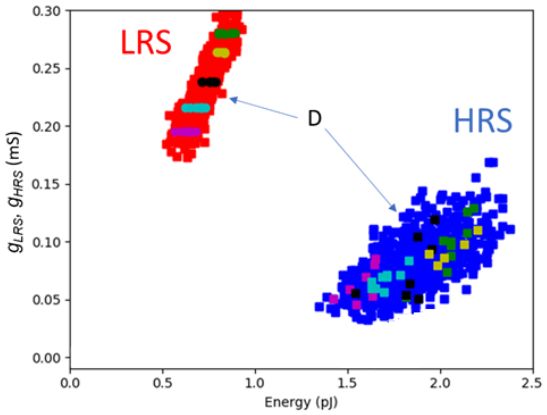


Figure 7. *Cycle-to-cycle distributions of conductivity in LRS and HRS states vs. energy consumed in transitions (SET and ReSET, accordingly) to these states for the device D. Five groups of ReSET transitions (not consecutive) from different LRS are color-marked. After ReSETs the initially tight LRS distributions in each group transform into rather wide HRS distributions (marked by the same colors).*

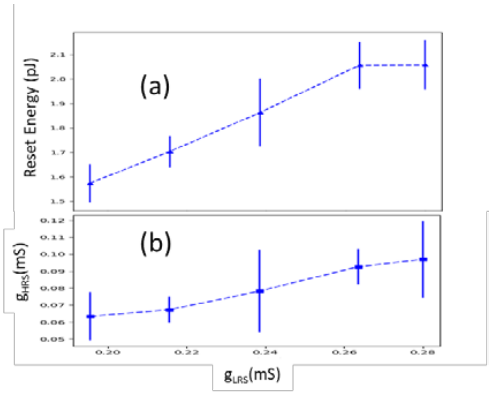


Figure 8. *(a) Switching energy in the ReSET processes and (b) conductivity in HRS in switching cycles marked by colored symbols in Fig. 7. Bars show the standards deviation of measured data.*

To outline the role of the starting state, Δg_{LRS} and Δg_{HRS} data of the device D in Fig. 7 are plotted vs. conductivity of the starting (pre- switching) states, from which each switching took place (Figs. 10, 12). In SET, consumed energy E_{Switch} is primarily determined by the final switching state since the LRS conductance dominates – therefore, larger starting resistance (larger g_{HRS}) should result in smaller Δg_{LRS} for a fixed E_{Switch} – in agreement with the data in Fig. 10. At the same time, Δg_{LRS} exhibits strong energy dependency as indicated by a larger slope of the energy gradient (normal to the equal-energy regions marked by the parallel darts lines) in Fig. 10. The trends of resulted LRS conductance and switching energy E_{Switch} correlate, Fig. 11. In SET process, the effect of variation of conductance of starting state, g_{HRS}, is weaker than the effect of the switching energies, as seen in Δg_{LRS} dependency on the slope of the dart line in Fig. 10.

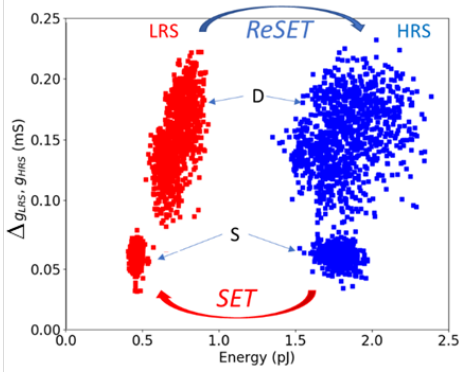


Figure 9. Distributions of the relative conductance changes, $\Delta g_{LRS} = g_{LRS} - g_{HRS}$ (SET) and $\Delta g_{HRS} = |g_{HRS} - g_{LRS}|$ (ReSET), vs corresponding switching energies for each cycle in D and S devices. Y-axis corresponds to the conductancies resulted from SET and ReSET switching, while x-axis contains values of energies released during switching.

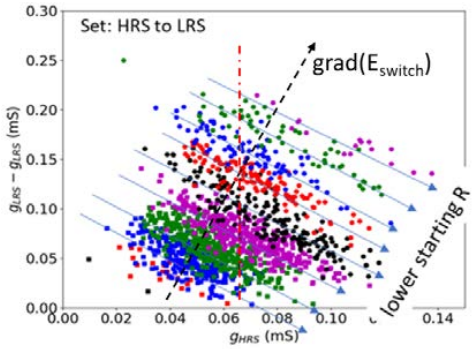


Figure 10. *Conductivity changes in HRS to LRS switching (device D), Δg_{LRS}, vs. starting g_{HRS} values. Symbols of the same color and type (along each individual dart line) identify switching operations occurred with same energy.*

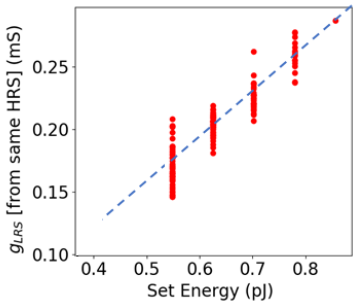


Figure 11. *Post-switching LRS conductance vs. energy for switching from the same HRS state as marked by the vertical dash-dot line in Fig.* 10.

Veksler, Dmitry; Bersuker, Gennadi; Bushmaker, A; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"Switching variability factors in compliance-free metal oxide RRAM."
Paper presented at 2019 International Reliability Physics Symposium, Monterey, CA, United States. March 31, 2019 - April 4, 2019.

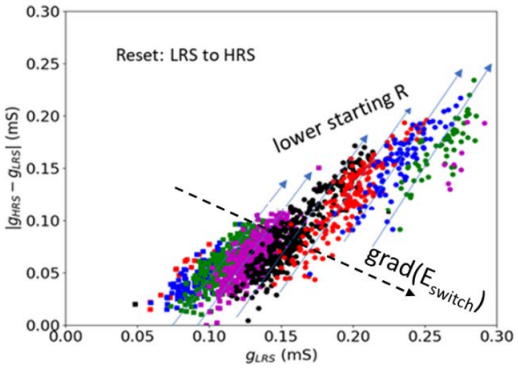


Figure 12. *Conductivity changes in Reset switching, $\Delta g_{HRS} = |g_{HRS} - g_{LRS}|$ vs. starting LRS conductance values. Symbols of the same color (along each individual dart line) identify switching operations occurred with same energy.*

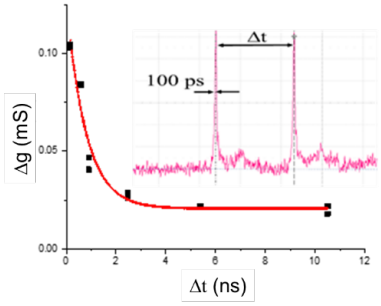


Figure 13. *Correlated effect of repeated pulses. Relative increase of RRAM conductance, Δg_{HRS} vs. time delay, Δt, between 2 sequential set pulses (see inset) when $\Delta t \leq 2$ns.*

In ReSET, an opposite trend is expected: since switching energy E_{Switch} is dominated by LRS, larger starting g_{LRS} should lead to larger Δg_{HRS}, as indeed is observed in Fig. 12. Thus, HRS resistance is primarily determined by the starting LRS value: consequently, lower starting resistance leads to greater switching window. The corresponding energy gradient is smaller than that of the starting resistances (as indicated by arrows in Fig. 12). In ReSET, resistance change is driven by the switching energy, which is controlled by LRS – higher g_{LRS} results in higher g_{HRS}. In SET, initial the beginning of the switching process, resistance change is driven by the consumed HRS energy, which then increases along with the increase of g_{LRS}, resulting in a weaker Set energy dependency/distribution, as seen in Fig. 7.

We speculate that LRS variability results in variations of the consumed energy (E_{Switch}) and temperature in the vicinity of the filament (depends on LRS current) during switching that strongly affects redox processes [4], and, in turn, may cause Reset cycling instability.

To assess the effect of temperature on switching, we studied how RRAM resistance is modulated by the timing between two sequential programming pulses (Fig. 13). Reducing the time interval between pulses below 2 ns results in significant amplification of their effect on resistance. It's consistent with theoretical expectations that, by increasing local temperature around the conductive path [1], the first pulse magnifies the structural changes induced by the subsequent pulse. The filament resistance in hafnia-based RRAM can be gradually and linearly changed depending on polarity of programming pulse, Fig 14 (the corresponding switching energy was estimated to be in sub pJ range, Fig. 15), that is an essential feature of an artificial synapse.

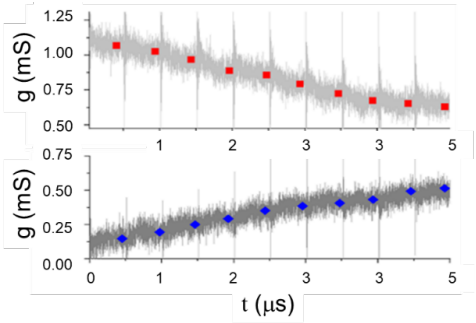


Figure 14. *Semi-linear increase and decrease of the RRAM conductance under continuous SET (0.75 V) and RESET (1.25 V) pulses. Program pulse width is 100 ps. Square marks on the conductance traces correspond to averaged values of the RRAM conductance after each pulse.*

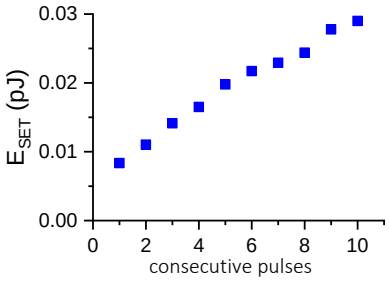


Figure 15. *Upper limit estimate of the energy consumed in each SET pulse operation in Fig. 14.*

IV. Conclusion

This study indicates that RRAM evaluation should be done under the frequencies and voltages conditions close to intended circuitry applications. In the sub-ns operation time range, hafnia-based devices demonstrate compliance-free forming and ultra-low energy switching.

Correlation between switching variability and energy consumed during switching process points to the effect of random structural changes in a certain (likely minor) region of the conductive filament. Higher filament resistance leads to lower dissipated energy: resulted lower temperature slows redox kinetic that is expected to suppress variability. Such property of hafnia-based devices facilitates their use in large scale neuromorphic cross-bar architectures.

Veksler, Dmitry; Bersuker, Gennadi; Bushmaker, A; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
"Switching variability factors in compliance-free metal oxide RRAM."
Paper presented at 2019 International Reliability Physics Symposium, Monterey, CA, United States. March 31, 2019 - April 4, 2019.

References:

1. G. Bersuker, D. Gilmer, D. Veksler “Metal oxide resistive random-access memory (RRAM) technology” *Advances in Non-volatile Memory and Storage Technology*, Elsevier (2019).
2. P. R. Shrestha, D. Nminibapiel, J. H. Kim, J. P. Campbell, K. P. Cheung, S. Deora, G. Bersuker, and H. Baumgart, Energy control paradigm for compliance-free reliable operation of RRAM, in proc. 2014 IEEE International Reliability Physics Symposium, MY.10.1-MY.10.4 (2014)
3. D. M. Nminibapiel, D. Veksler, P. R. Shrestha, Ji-H. Kim, J. P. Campbell. J. T. Ryan, H. Baumgart, and K. P. Cheung, "Characteristics of Resistive Memory Read Fluctuations in Endurance Cycling," in *IEEE Electron Device Letters*, **38**, pp. 326-329 (2017).
4. G. Bersuker, D. Veksler, D. M, Nminibapiel, P. R Shrestha, J. P Campbell, J. T. Ryan, H. Baumgart, M. S. Mason, K. P. Cheung, *Journal of Comp. Electronics*, **16**, 1085-1094 (2017).

Veksler, Dmitry; Bersuker, Gennadi; Bushmaker, A; Shrestha, Pragya; Cheung, Kin; Campbell, Jason.
”Switching variability factors in compliance-free metal oxide RRAM.”
Paper presented at 2019 International Reliability Physics Symposium, Monterey, CA, United States. March 31, 2019 - April 4, 2019.

Sources of Errors in Structured Light 3D Scanners

Prem Rachakonda, Bala Muralikrishnan, Daniel Sawyer
Dimensional Metrology Group,
Sensor Science Division,
National Institute of Standards and Technology, Gaithersburg, MD

1 INTRODUCTION

The Dimensional Metrology Group (DMG) at the National Institute of Standards and Technology (NIST) has been involved in the development of documentary standards for a variety of 3D metrology instruments such as coordinate measuring machines (CMMs), laser trackers and terrestrial laser scanners (TLSs). As a US National Metrology Institute (NMI), NIST develops procedures to evaluate dimensional metrology instruments in an unbiased and objective manner. Lately, several of DMG's industry partners have shown interest in Structured Light (SL) scanners and have expressed the need to characterize the instrument performance.

Structured light scanners are portable short-range 3D imaging systems that use patterns of projected images or light patterns to measure an object. These scanners have very few or no moving components and capture millions of points on surfaces of an object in their field of view. SL scanners build on classical photogrammetry techniques and project encoded images on an object, which add texture to the object and simplify the correspondence problem, i.e., corresponding points in the projected and the captured images. The precision and accuracy of measurements from SL scanners can be appropriate for certain industrial applications such as reverse engineering, quality control, and biometrics. These scanners also have faster data acquisition rates and are relatively inexpensive compared to the large stationary CMMs. This makes them suitable for applications that do not need the low uncertainties typical of CMMs or those that require in situ measurements.

SL scanners have been commercially available for over a decade and some commercial scanners are evaluated using one of two German guidelines – VDI/VDE* 2634 parts 2 and/or 3 [1,2]. Several other research groups and NMIs have developed physical artifacts that are agnostic to instrument construction and are purpose driven. The use of such guidelines and artifacts is not well understood for instruments which have a variety of sensor configurations, projected patterns, sensor/work volumes, point densities, triangulation angles and targets of varying size & form. It is also not clear if these guidelines/artifacts are sensitive to all the sources of errors that are present in these systems. The two VDI/VDE 2634 guidelines for evaluating the performance of the SL scanners use artifacts of known mechanical and optical characteristics, but the real-world usage of these scanners may involve objects of varying characteristics. The end user may not be able to make informed decisions if an instrument is specified based on these guidelines/standards but is used for an application that deviates from the test conditions. This can cause an enormous financial burden for organizations that intend to invest in the technology, but do not have appropriate standards to aid in selecting an instrument that meets their needs.

In this context, this paper will describe the ongoing activities at NIST to study various sources of errors in SL scanners with an objective of characterizing their performance. The paper will first give some background information on the principle of operation of SL scanners, the state-of-the-art of documentary standards/artifacts and describe some scanners available to the authors. The paper will then discuss the various sources of errors in SL scanners that influence their

* German: Verein Deutscher Ingenieure/Verband der Elektrotechnik (English: Association of German Engineers/Association of Electrical Engineering)

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

measurements and describe a few experiments that were performed to understand some of these errors.

2 PRINCIPLE OF OPERATION

Structured light scanners are systems that acquire 3D coordinates of points on surfaces of an object using the principle of triangulation. These scanners project a pattern of light on an object, capture the distorted patterns from the object using a camera and calculate the 3D coordinates. The principle of triangulation used to calculate the depth of the projected points on the object is based on the knowledge of the distance between the pattern projector and the camera and the angle between them with respect to a location on the object.

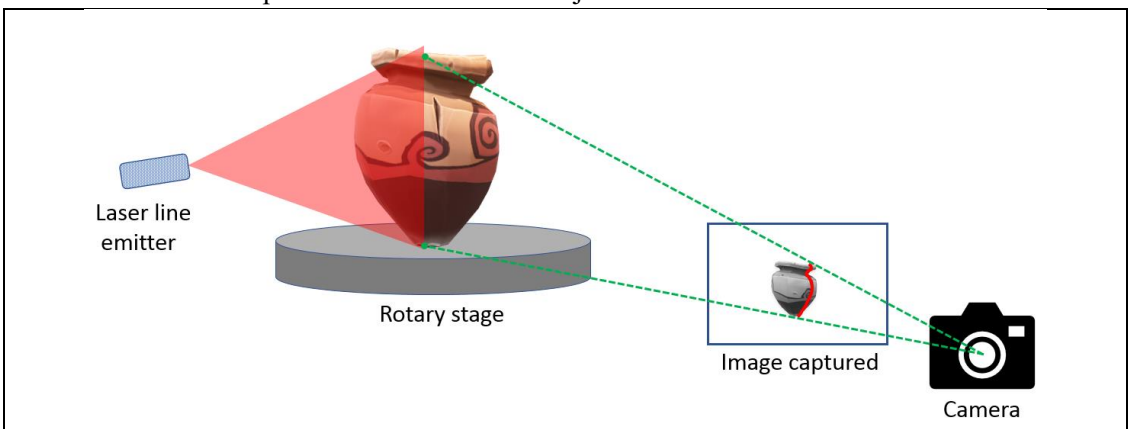


Figure 1: Illustration of a laser line scanner [3]

To explain the principle of operation of an SL scanner, we will first explain the principle of a laser line scanner, which also uses the principle of triangulation. A laser line scanner has a laser line emitter that projects a plane of light on to an object, and a camera that captures the distorted image of the line corresponding to the object's contour (see Figure 1). To obtain a complete 3D scan, such a scanner needs the object to move relative to the laser line. For example, the object can be mounted on a linear or rotary stage and moved relative to the scanner or an opto-mechanical system may be used to sweep the laser line over the object to obtain a 3D scan.

A structured light scanner uses the same principle, but instead of a laser line emitter, it uses a projector to illuminate the object (see Figure 2). The purpose of the projector is to "paint" the object with a known pattern that can be easily captured using a camera. However, to detect these patterns from the image, common edge/line detection algorithms are not enough. Any illumination change will result in loss of data or detection of lines at a wrong location. To address this issue, a structured light scanner uses a projector to display lines in several patterns or colors (structured light) to "paint" the object. There are several coding techniques for reducing the error of detecting the lines from the captured images. This is possible because modern projectors are capable of projecting 1920 lines for a high-

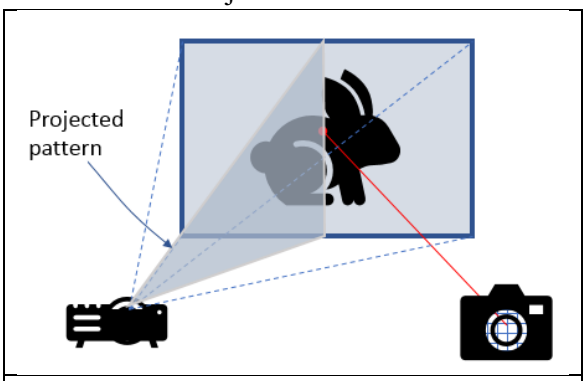


Figure 2: Illustration of an SL scanner that uses a plane-line intersection

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

definition (HD) projector and 3840 lines for a 4K projector, where the number of lines correspond to the projector's columns in pixels.

The purpose of a coded light pattern/image is to easily differentiate between two lines in the distorted image as captured by the camera. For example, the first column of the projected image could be red, and the next column could be yellow and so on. And, like a laser line scanner, the triangulation principle is used to recover the depth information.

While discussing structured light scanners, it is important to mention one other widely used technique – photogrammetry, and specifically stereophotogrammetry, which also uses the triangulation principle. Photogrammetry based instruments are considered passive vision imaging systems that rely on the distinct features in the scene to obtain the correspondence between the images in two or more cameras or multiple images from the same camera from multiple positions. When such features do not exist, the instrument fails to perform the triangulation adequately. For example, measuring the form of the painted exterior of an automobile with a slight dent will be challenging using photogrammetry [4]. On the other hand, structured light scanners are considered active vision scanners and use known image patterns to achieve correspondence between the projected image and the captured image.

3 STATE-OF-THE-ART OF PERFORMANCE EVALUATION STANDARDS

Common specifications and/or common procedures to evaluate scanners will help the end users and decision makers in choosing the right scanner for their application. Scanners may be evaluated in many ways, one of which is to develop a set of procedures that use artifacts of common geometry such as planes, spheres and cones. Another way is to develop specific reference artifacts with complex geometry and corresponding procedures that encompass the application requirements. Whichever method is used to characterize a scanner, it should be acceptable to both users and manufacturers of SL scanners. The next sub-sections will discuss some of the existing methods to characterize optical 3D scanners such as SL scanners.

3.1 Reference artifacts

The National Research Council of Canada (NRC) and the National Physical Laboratory of United Kingdom (NPL) are the NMIs of their respective countries and have been working for several years on developing artifacts [5,6,7] shown in Figure 3 and Figure 4 for short-range optical 3D scanners. Though the use of these artifacts has not been adopted yet by any documentary

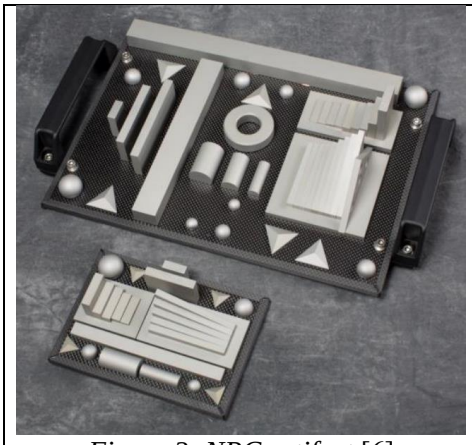


Figure 3: NRC artifact [6]

Figure 4: NPL artifact [7]

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

standard, they provide a way for end-users to characterize their instrument. Many organizations also develop task specific artifacts and/or procedures to address their internal needs.

3.2 Documentary standards

As of the writing of this paper, there are no international documentary standards for SL scanners. There are two German guidelines[†] that are used to evaluate some SL scanners, namely the VDI/VDE 2634 Parts 2 and 3. These two guidelines were written as an extension of the International Standards Organization (ISO) 10360 standards for CMMs, instruments whose operating principles are much different than those of SL scanners [6]. In 2018, the ISO Technical Committee (TC) 213 initiated standards development work on a Part 13 to the ISO 10360 series that addresses the evaluation of SL scanners. Thus far, this part includes some of the procedures of the VDI/VDE 2634 guidelines, along with more exhaustive testing on the number of points and point filtering criteria. Since this document is still under development, its discussion is out of the scope of this paper.

Table 1: Comparison of the two VDI/VDE 2634 guidelines

	VDI-VDE 2634 Part 2 (2012)	**VDI-VDE 2634 Part 3 (2008)**
Multiple views	No	Yes
Filtering of points	Limited specification on point rejection	
	Probing errors	
Artifact	Sphere	Sphere
Sphere diameter*	$0.02L_0$ to $0.2L_0$	$0.02L_S$ to $0.2L_S$
Positions	10	>=3
Scans/position	1	>=5
Quality parameter	Error in sphere form	Error in sphere form
	Error in sphere diameter	Error in sphere diameter
	Sphere spacing errors	
Artifact	Like a dumbbell	Like a dumbbell
Sphere diameter*	$0.02L_0$ to $0.2L_0$	$0.02L_S$ to $0.2L_S$
Dumbbell length*	$>=0.3L_0$	Varies
Positions	7	7
Scans/position	1	1
Quality parameter	Error in sphere-sphere distance	Error in sphere-sphere distance
	Flatness measurement error	**Length measurement error**
Artifact	Flat plate	Ball bar/ball plate/gage blocks
Artifact width	Minimum 50 mm	N/A
Artifact length*	Minimum $0.5*L_0$	$0.02L_S$ to $0.2L_S$
Positions	>=6	7
Scans/position	1	1
Quality parameter	Flatness measurement error	Error in artifact length
* A dumbbell can be replaced by two spheres separated by a rigid mount.		

[†] Though guidelines are informative, these documents are considered as *de facto* standards by the practitioners. There exists a VDI/VDE 2634 Part 1, but it is not applicable to imaging systems based on area scanning, like SL scanners.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

The two VDI/VDE guidelines mentioned here use simple geometries such as spheres in multiple positions and orientations to evaluate optical scanners. One of the major differences between the two VDI/VDE guidelines is that Part 2 is applicable for scanners that produce area scans based on a single view and Part 3 is applicable for scanners using multiple views that extends the measurement volume. Table 1 provides a more detailed comparison of the two guidelines. In this table, L_S is the length of sensor measuring volume diagonal and L_0 is the body/spatial diagonal of the measuring volume, both of which are specified by the manufacturer.

There are several areas where these two guidelines are not adequate for choosing SL scanners today. These two guidelines do not address all the scanner configurations that are currently available, nor provide all the parameters that describe the performance of a scanner. Below are some of the issues in these two guidelines that could be improved.

a) The VDI/VDE guidelines do not recommend any tests to determine the scanner's resolution in the X, Y, Z axes for a user to determine the minimum size of a feature that can be measured by the scanner. Some scanners offer a theoretical lateral resolution (X, Y) based on the camera sensor's pixels and the field of view, which does not indicate the smallest feature that can be distinguished.
b) VDI/VDE 2634 Part 3 does not require the calculation of the flatness measurement error, whereas it is required by Part 2.
c) Accuracy values of these instruments vary based on the location of the artifact and/or distance of the artifact from the scanner. A single accuracy value may not be adequate for the end user.
d) Recommended artifact geometries are either flat or convex (spheres). The performance of some of these scanners was found to be different for concave surfaces, especially when scanning deep, dark and/or shiny features.
e) The guidelines are unclear for merged 3D scans obtained using different scanner settings to capture the surface without any relative displacement between the scanner and the object. Such scans are typically needed when the object is dark and/or shiny and are obtained by merging scans at various exposure, gain or aperture settings.
f) The guidelines are also not clear when the scanner uses multiple cameras and generates scan data by merging the scans from multiple cameras.
g) Both the VDI/VDE guidelines do not have a specification on the number of points to consider. Point densities affect the error in calculating the quality parameters.
h) Both the VDI/VDE guidelines have criteria for rejecting a certain percentage of points, but do not have a prescribed method to reject those points.

4 STRUCTURED LIGHT SCANNERS

There are several configurations of structured light scanners that are commercially available, and a discussion of the various scanners and their features is beyond the scope of this paper. Some of these scanners are priced from $100 to $200K, with various hardware & software options and performance characteristics. The price of the instruments may not correlate well with their performance, but it is important to note that the same principle is used in products for various applications at various price ranges. SL scanners may use different coding patterns, hardware, filters and proprietary algorithms. DMG procured three commercial structured light scanners to study their construction and performance.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

Table 2: Types of scanners used by DMG				
Name	**SK**	**SE**	**SA**	**SC**
Application	Gaming	Hobby	Industrial	Experimental
# of cameras	1	1	2	1
Camera Resolution	0.3MP	1.3MP	8MP	2MP
Projector type	LED pattern	LED lamp	LED lamp	Metal halide lamp
Light color	Infrared	White	Blue	Multi-color
Light pattern	Pseudo random binary dots	Hybrid	Gray/binary code	Gray code
Projector resolution	N/A	<2 MP	28MP	2 MP
Single scan volume	≈ 2 m^3	≈ 1.73 m^3	≈ 0.11 m^3	≈ 2 m^3
Stated accuracy	NA	<0.050 mm	0.032 mm	N/A

There are also several implementations of structured light scanners in open literature including open source software. DMG researchers built an SL scanner (scanner SC) based on the design and software described by Taubin et al [8] that uses a plane-line intersection as illustrated in Figure 2. Some of the specifications of these four scanners are given in Table 2 and the scanner SC will be described next.

4.1 Custom-built 3D scanner

The purpose of building a 3D structured light scanner was to enable the authors to modify all the parameters that affect the instrument performance – a capability that was limited by commercial instruments. Even though commercial scanners resulted in higher quality scans than this custom scanner, the ability to modify the instrument parameters is very useful in understanding the instrument error sources.

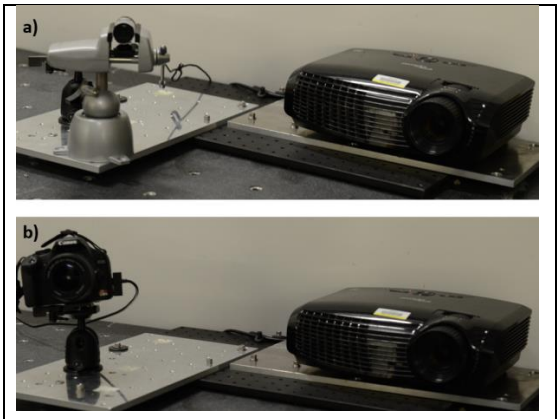


Figure 5: Custom-built SLS

The hardware for the custom-built SL scanner includes a commercial off-the-shelf projector (Optoma HD 143X), a webcam (Microsoft LifeCam Studio) both of which have a resolution of 1920 pixels × 1080 pixels and shown in Figure 5a. The software was a pre-compiled binary of the C++ code developed by Taubin et al [8]. This scanner uses a series of temporal gray code patterns as shown in Figure 6. The software also has a projector-camera calibration routine and a scanning routine that generates 3D point cloud data. Figure 6 and Figure 7 show a picture of the artifacts being scanned and the 3D data respectively using the webcam.

After this initial build, the webcam was replaced by a digital single lens reflex (DSLR) camera (Canon T1i, shown in Figure 5b) to improve the performance of the scanner. A third-party commercial software was used to convert the DSLR to a UVC (USB video class) compatible camera. This additional software enabled a simpler software integration, instead of modifying the C++ code that only worked with UVC webcams. The DSLR camera provided a way to control the camera parameters much more easily than what a consumer grade webcam would allow. It also allowed the usage of a higher resolution images, larger sensor size and higher quality lenses

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

resulting in images with higher signal-to-noise ratios. It should be noted that higher camera pixel resolution is different from the resolving power of the lens, i.e., its ability to differentiate two lines as separate.

5 PARAMETERS THAT AFFECT SCANNER PERFORMANCE (ERROR SOURCES)

Measurements obtained by structured light scanners are affected by several parameters. These parameters are the sources of errors in these scanners. In commercial instruments, many of the error sources are addressed using higher quality hardware, optics, rigid mounting, and hardware/software filters. Error sources can be categorized as caused by a) scanner construction, b) target construction, c) environment, d) operating modes/settings and e) data acquisition and post-processing. These error sources are described in detail in the following sub-sections.

5.1 Errors due to scanner construction

5.1.1 Calibration (intrinsic and extrinsic) and distortion parameters

SL scanners need to be calibrated before they can be used for measurement. This calibration process typically uses an artifact with a grid of known 2D patterns and will result in the calculation of calibration parameters. The methods used in performing calibration are described next.

a) Camera calibration

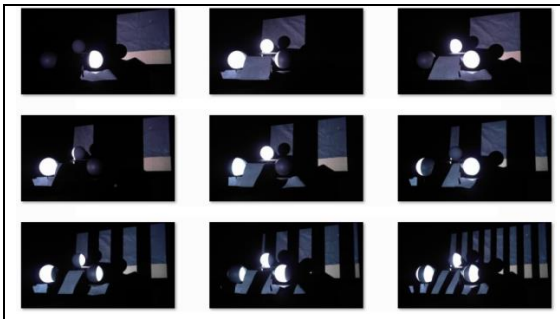


Figure 6: The first nine sequences of the gray code pattern used by the custom-built SLS

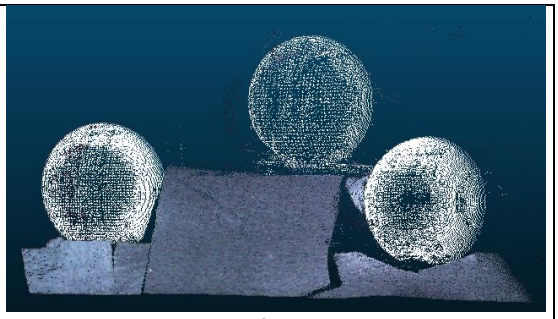


Figure 7: 3D scan of three spheres scanned using a custom-built SLS

Camera calibration has been studied extensively [9,10,11,12] and is well understood in the field of computer vision. It is described in this paper in brief because the parameters that affect camera calibration also affect the structured light scanner calibration. Cameras used for SL scanners are calibrated based on a general projective camera model [11] and it is a generalized model of a pinhole camera. It requires the determination of five intrinsic parameters and six extrinsic parameters of a camera.

The calibration process based on this model is valid only for undistorted images, as a pinhole camera does not use any lens. So, the first step in this calibration process is to generate an undistorted image by minimizing the radial and tangential distortions due to the camera lens illustrated in Figure 9 and Figure 10 respectively. This involves the calculation of typically three parameters (k_1, k_2, k_3) for radial distortion and two parameters (p_1, p_2) for tangential distortion[‡].

[‡] The calibration routine described in this paper, MATLAB and the open source computer vision library (OpenCV) use different notations for the matrices and the distortion parameters. The parameters described here are per OpenCV.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

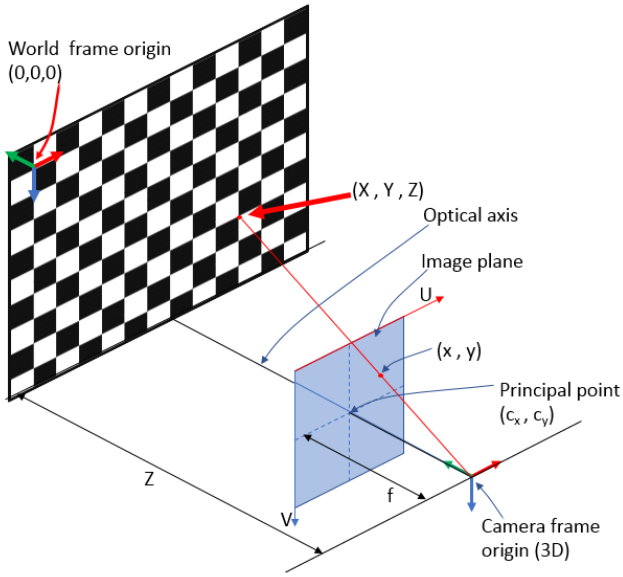


Figure 8: Pinhole camera model

The relationship between the 2D image points captured by the camera and the 3D world points is given by equation 1. Here X represents the world coordinates of a point $[X, Y, Z,1]^T$, x represents the undistorted image coordinates $[u, v,1]^T$, R is a 3×3 rotation matrix, t is a 3×1 translation matrix and K represents a 3×3 intrinsic matrix of a camera. The rotation and translation matrices represent the relationship between the origin of the world points (typically set arbitrarily on the calibration artifact) and the camera's origin as shown in Figure 8.

The camera's intrinsic matrix K is given by equation 2, which is composed of five intrinsic parameters. These are, the focal lengths f_x, f_y, the optical center or principal point in two orthogonal directions c_x, c_y and axis skew s.

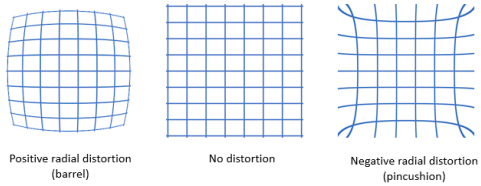


Figure 9: Effects of radial distortions

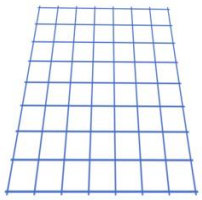


Figure 10: Effects of tangential distortion

$$x = K[R\ t]X. \qquad 1$$

The skew coefficient (s) is to account for any non-orthogonality in the image axes (non-rectangular pixels). Typically, a pinhole camera model doesn't have this issue of skew but can be a result of certain digital conversion operations. The two focal lengths f_x, f_y are to account for rectangular pixels and are calculated from the physical focal length (F) in mm and the size of pixels m_x, m_y respectively in the units of pixels/mm [11]. That is, $f_x = F \times m_x$ and $f_y = F \times m_y$.

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \qquad 2$$

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

The extrinsic camera parameters have six degrees of freedom that describe the position of the camera in the "world" (three translations and three rotations) and the intrinsic parameters K has five degrees of freedom, a total of 11 degrees of freedom. Equation 1 describes the correspondence between 3D world points X and 2D pixel locations of an undistorted image x, which can also be written as equation 3:

$$x = PX, \qquad 3$$

where $P = K[R\ t]$ and is called the "camera matrix". For a general projective camera, the camera matrix P is given by equation 4, where p_{34} is always equal to 1. Therefore, P has 11 degrees of freedom and a matrix rank of 3[11].

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}. \qquad 4$$

b) <u>Camera-projector calibration</u>

There are many commercial and open source software that can achieve a single or stereo camera calibration, however there are only a handful of research efforts that describe the calibration of a camera and a projector [13,14,15]. These research efforts throw light on the ways a scanner manufacturer may approach the process of calibrating an instrument. We briefly describe one such method to perform a camera-projector calibration.

The primary reason to perform the camera-projector/scanner calibration is to reduce the average reprojection error. Reprojection error is defined as the average of the distance between points projected to the camera's image plane using the camera parameters (intrinsic, extrinsic and distortion) and the points captured by the camera. It is the correction needed to obtain a near-perfect correspondence between the projected and the captured images [9]. This reprojection error convolves all the intrinsic, extrinsic, radial, and tangential distortion parameters of both the projector and the camera. Lower reprojection errors indicate a better estimate of the camera parameters. Once calibrated, the intrinsic parameters are typically constant for an SL scanner. This assumes that the projector's focus and magnification settings are constant, and the camera's settings for focal length and aperture are constant. The camera's settings can be maintained constant by disabling any automatic settings and enabling a manual mode.

The extrinsic parameters of an SL scanner can change if the relative position or orientation between the camera and the projector is changed due to intentional or accidental mechanical movement. To simplify the calculations in our implementation, once the SL scanner was calibrated, the "world" origin was arbitrarily set to either the camera's origin or the projector's origin. In this case, the rotation and translation parameters then referred to those between the camera and the projector. The extrinsic parameters between the camera and the projector can be maintained constant by using rigid mounting apparatus for both the camera and the projector. If these conditions are not met, the system must be recalibrated before every scan.

The custom-built SL scanner (SC) uses a calibration routine [14] and needs at least three sets of data, with each dataset consisting of multiple images of the grid plate in different positions that are used for camera calibration first and then the scanner calibration. More positions will lower the reprojection error. To understand the sources that dominate the reprojection error, 14 calibration datasets were obtained and three datasets at a time were processed using scanner SC's calibration routine. The reprojection errors of the camera, projector, and the scanner (stereo) are shown in Figure 11a. It may be observed that the stereo component dominates the camera or the projector reprojection error. Both the camera and projector reprojection errors have low mean values and low variability compared to the stereo reprojection errors. When the same process was

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

repeated, now using 13 datasets at a time, the stereo calibration error reduced considerably and is shown in Figure 11b.

The intrinsic parameters from the calibration are given in Table 3. The units of f_x, f_y, c_x, c_y, s and the reprojection error (e_R) are pixels and k_1, k_2, k_3, p_1, p_2 are the radial and tangential distortion parameters that are dimensionless. Only 75 dataset combinations of the possible 364 ($^{14}C_3$) were considered to calculate the statistics for Table 3a and the calibration routine failed to calculate the parameters for one of the 14 possible combinations in Table 3b. From both Table 3 and Figure 11, it is evident that using more calibration datasets is beneficial in lowering the bias and variation in the reprojection errors, but also that the projector's intrinsic parameters have higher bias and variation than that of the camera. This is to be expected since the projector's optics are typically of lower quality than that of a camera.

Table 3: One standard deviation values of the intrinsic, extrinsic and distortion parameters of the custom-built scanner

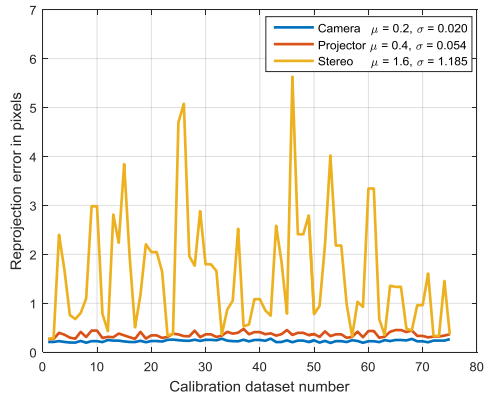
	Camera	Projector
f_x	46.685	154.455
f_y	39.196	117.146
c_x	4.375	32.348
c_y	41.470	127.601
s	0.000	0.000
k_1	0.009	0.093
k_2	0.012	1.304
p_1	0.001	0.006
p_2	0.002	0.004
k_3	0.000	0.000
e_R	0.020	0.054

a) Calibration using 3 datasets at a time

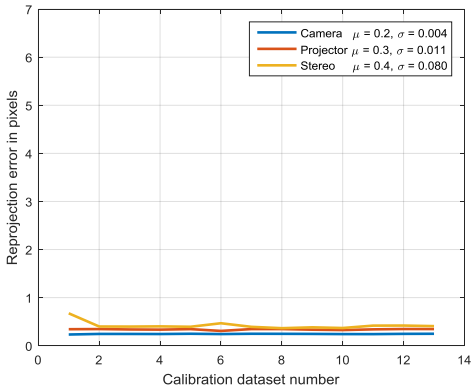
	Camera	Projector
f_x	1.633	13.463
f_y	1.411	9.448
c_x	0.905	2.754
c_y	1.552	18.195
s	0.000	0.000
k_1	0.001	0.005
k_2	0.003	0.007
p_1	0.000	0.000
p_2	0.000	0.001
k_3	0.000	0.000
e_R	0.004	0.011

b) Calibration using 13 datasets at a time

The reprojection error affects the measurements such as sphere-to-sphere distance etc. A reprojection error less than 0.1 pixel is considered acceptable in practice, however, obtaining a correlation between the reprojection error and error in measuring an artifact (say a dumbbell) is not trivial. This is because, larger reprojection errors distort the 3D point cloud data to such an



a) Three datasets at a time



b) 13 datasets at a time

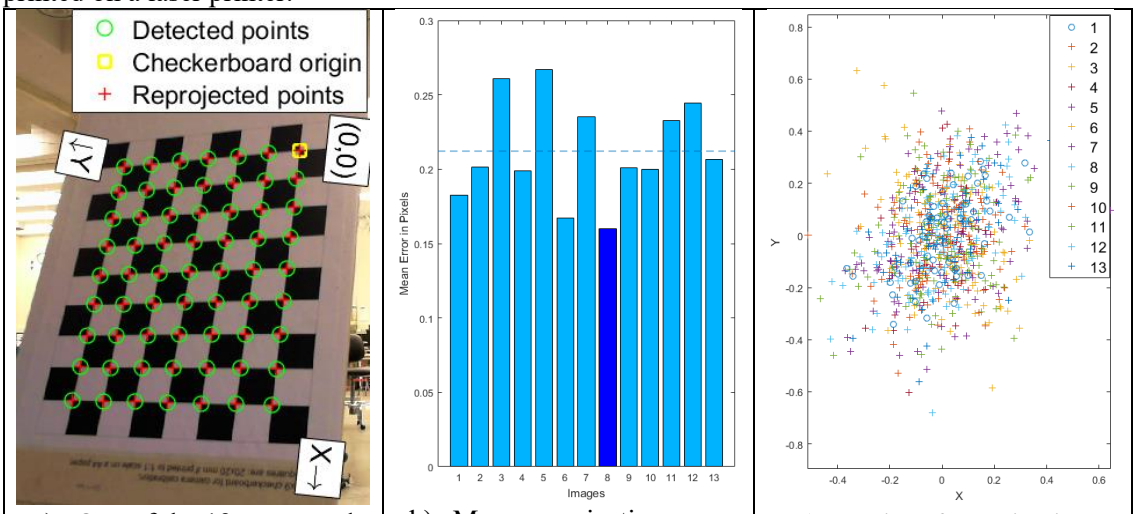
Figure 11: Reprojection errors of the camera, projector, and stereo

extent that the scan of a sphere does not appear to be a sphere. However, the distance (d_{CP}) between the camera and the projector of the custom-built scanner (SC) was calculated from the translation component of the extrinsic parameters (stereo). This distance (d_{CP}), correlated well with the reprojection error (e_R) in multiple calibration routines and the correlation coefficient was 0.85.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

5.1.2 *Calibration artifact*

The calibration artifact recommended by many software suites is a pattern of images on a flat plate. These patterns could be anything that can be easily detected by the image processing software and are typically 2D checkerboard patterns or circular patterns. These patterns can be easily printed on a paper using a laser printer and adhered to a rigid plate. However, most software assume that the grid pattern is uniform, and that the flatness of the pattern is zero. If the flatness is large, or if the laser printer has a scaling/skew error, it will increase the reprojection error. Figure 12 shows an example of the reprojection errors calculated by MATLAB using a calibration pattern printed on a laser printer.



a) One of the 13 processed checkerboard images

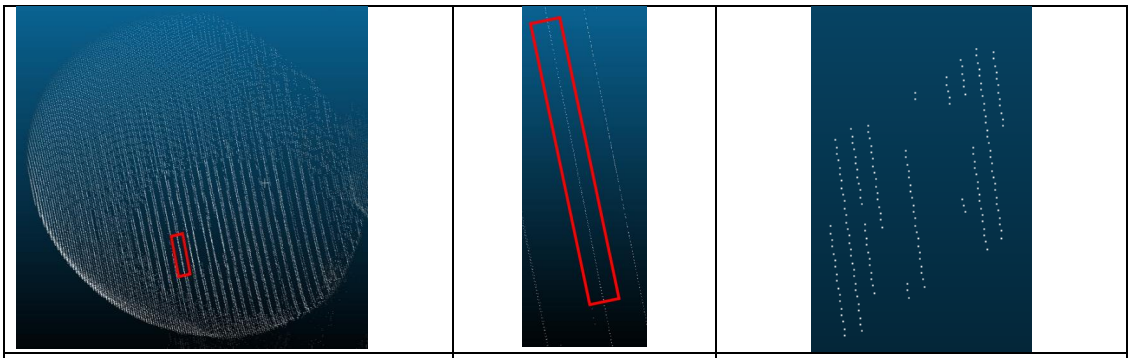
b) Mean reprojection error in each image

c) A plot of reprojection errors in 13 images

Figure 12: An example of reprojection errors calculated using MATLAB

5.1.3 *Camera and projector resolution*

In an ideal scanner, a single coded line projected by a projector is imaged by a camera and the line's image is exactly one pixel wide. Alternatively, the projected line in the structured pattern itself may be intentionally wider. However, when the pixel resolution of the projector and camera vary, it might result in redundant points. These points are typically away from the object's surface along the line joining the camera and the pattern's projection location on the object.



a. 3D scan data of a sphere

b. cropped data

c. side view of cropped data

Figure 13: Close-up view of the 3D points obtained using a custom-built SLS

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

To explain this effect, let us assume that a single line projected by a projector is captured by a camera as a distorted line that is 10 pixels wide. If the method of calculation of the 3D point involves a plane-point intersection, it will result in 10 points for a single coded line from the projector. These points must be filtered to get an accurate representation of the object. Figure 13 shows this issue using data obtained from a 100 mm diameter sphere at 0.67 m from scanner SC. Figure 13c shows the side view of the data in Figure 13b which appear to be penetrating the sphere surface. The missing points in Figure 13c could be a result of the lack of enough intensity of the pattern in the captured image. The points in Figure 13c were fit to a plane and the one standard deviation (1σ) value of the residuals was 5.8 nm – a negligible value. This indicates that these points are indeed redundant and may be filtered out.

5.1.4 Camera and projector's depth of field

Commercial DSLR cameras have variable apertures that enable a user to set the working distance and the depth of field (DOF), which is the distance between the nearest and the furthest part of an image that appears to be sharp. Knowledge of DOF is necessary to determine the scanning volume of an SL scanner. Larger DOF ensures that the images of the patterns captured are sharp resulting in lower noise. Figure 14 shows the variation of the DOF with varying aperture diameter for one DSLR camera.

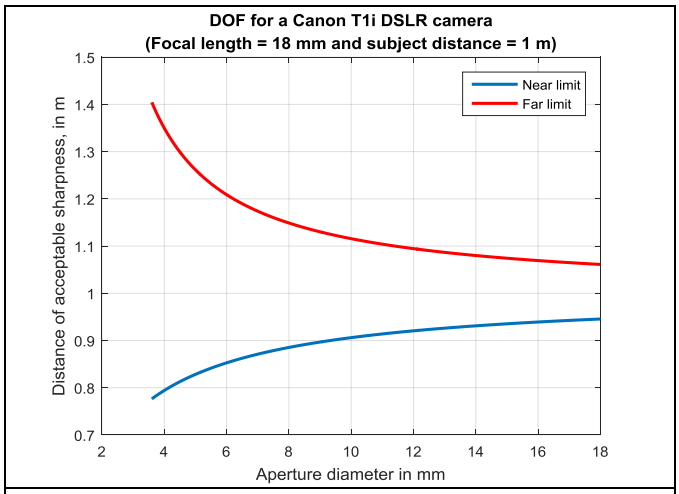


Figure 14: DOF variation with aperture diameter

Projectors on the other hand have a very narrow DOF that cannot be adjusted, which limits the ability of the projector to project patterns with sufficient sharpness on the objects. Commercial projectors are designed to output the maximum amount of light without any hinderance and therefore do not have any aperture control and consequently have limited DOF.

The projector used for scanner SC has a narrow DOF of ≈ 50 mm that was measured by visual observation of projected image sharpness, whereas the artifacts measured using the scanner are of 100 mm in diameter, placed at distances over 100 mm from one another. There are several methods to enhance the DOF of such projectors in open literature. One such method is described by Iwai et al [16], that uses an electrical focus tunable lens (FTL) in front of projector's objective to increase its DOF. However, it is not clear if any of the commercial scanners are using any specific methods to enhance the DOF of projectors.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

5.1.5 *Triangulation angle/configuration*

Since the principle of an SL scanner is based on triangulation, the measurements are affected by the angle θ between the camera, the object and the projector (illustrated in Figure 15). The measurement is also affected by the distance between the camera and the projector, *b*, and the distance between the camera and the object, *d*. The relationship between the parameters in this setup is given by equation 5 [17]. When $\theta \approx 0°$, $\alpha + \beta \approx 180°$ and $b \approx 0$. This will result in large errors due to uncertainty in calculating the depth *d*. When $\theta \approx 90°$, the camera may not capture any useful image of the pattern and can be impractical when measuring deep features or holes. A typical arrangement involves keeping the angle $\theta \approx 30°$.

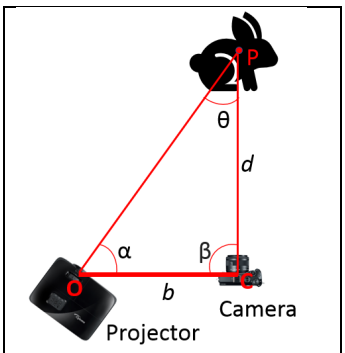


Figure 15: Configuration of an SL scanner

$$d = b\left(\frac{\sin(\alpha)}{\sin(\alpha+\beta)}\right) \qquad 5$$

5.1.6 *Coding techniques*

Geng [17] details a variety of techniques to code the light patterns, which include temporal (multi-shot/sequential) and spatial (single-shot) techniques. Multi-shot techniques are useful when the object being scanned is stationary, and results in more accurate measurements than single-shot techniques. Coding techniques affect the performance of the scanner [18] and each of these techniques have their specific applications, advantages, and disadvantages. Many scanners do not offer a way to change the light pattern coding and scanners are optimized for certain kinds of applications such as industrial metrology, gaming or dentistry.

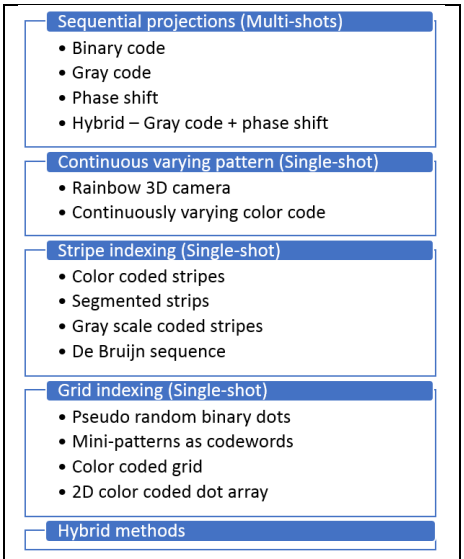


Figure 16: Various structured light coding patterns [17]

5.1.7 *Wavelength/Color of light for monochromatic coding patterns*

Some of the latest commercial structured light scanners use a blue light source. Some manufacturers claim that this type of light source reduces errors due to ambient lighting and reduces reflections due to short wavelength. Blue light is useful for some applications like scanning skin and teeth for dental applications [19]. While using blue light, the hardware and software allow the scanner to filter out ambient light and obtain higher quality scans. It is, however, not well understood if blue light scanners offer any significant advantage over white, or any other color lighting for most other applications.

5.1.8 *"Rainbow effect" in projectors*

Certain digital light processing (DLP) projectors use a mechanical color wheel to generate sequential color. Such a color wheel can interfere with the coding techniques described earlier due to a visible color artifact known as the "rainbow effect"[20]. Coupled with the camera's exposure

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

time variation, the rainbow effect can result in measurement errors. The mechanism to rotate the disc may also cause the projector to vibrate and can cause errors in measurements.

5.1.9 Lamp temperature

Many projectors use halogen or metal halide lamps as their light source and their operating temperatures can be anywhere from 35 °C to 300 °C and can cause the scanner components to thermally move with respect to each other. Most instruments have a warm-up period that addresses this issue. The projector's lamp is a major heat source that can also affect the mounting apparatus and thereby affect the calibration parameters of the scanner. Some scanners offer projectors with LED light sources which operate at lower temperatures that mitigate this issue.

5.2 Errors due to target/artifact shape and optical/mechanical properties

The VDI/VDE 2634 guidelines recommend the use of spheres and flat plates that have diffusely scattering surfaces. These kinds of targets are known to offer near-lambertian surfaces that provide data with low noise. However, such surfaces may not always be found in practical applications. Most optical scanners perform poorly with shiny and/or dark surfaces. Scanning concave surfaces also is challenging for scanners with large triangulation angles.

5.3 Errors due to environment

5.3.1 Ambient lighting

Ambient lighting is one of the major sources of error that can affect the quality of the data. The reconstruction algorithm of the SL scanner relies on observing the pattern in the captured image which may be coded based on light intensity from the projector. For example, consider a binary code used to digitally paint an object with a line numbered '23' whose binary representation is '10111'. If a speckle of light from the environment alters the intensity of light and saturates the image captured by the camera, it will result in an image that is coded '11111' which in decimal is '31'. This will cause an error in calculation of the depth using triangulation. However, in practice, it is typical to use intensity thresholding to exclude pixels that do not fall between certain preset intensities, which will result in missing points in 3D data. We also observed that some scanner manufacturers compensate for ambient lighting that mitigates this issue.

5.3.2 Ambient Temperature

Ambient temperature variation is an important source of error in many dimensional measurements. If the instrument is calibrated at one temperature and is used to measure an object at a different temperature, the measurement may be erroneous due to the lack of mechanical stability within the temperature range.

5.3.3 Other environmental factors

Factors such as humidity, fingerprints, smudges, and dust can result in the optics getting fogged up and/or dirty resulting in measurement errors.

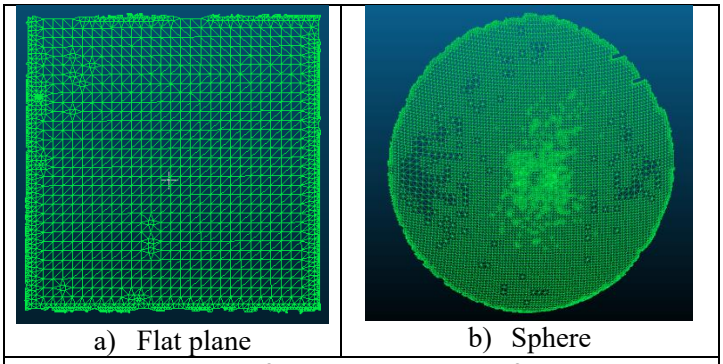
5.4 Errors due to mode of operation

SL scanners have several controllable parameters that are designed to generate a scan that is optimal for the target. The level of control to acquire scans varies with the instrument type and manufacturer. Parameters such as exposure time, gain, aperture diameter and exposure bracketing can result in datasets that have varying levels of noise. Not all scanner manufacturers offer a way to control these parameters, leading to the inability of an end user to compare scanners.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

5.5 Errors due to data acquisition and post processing

Errors in characterizing SL scanners can also occur due to the way the scanner acquires data and the methods used to process the acquired data. Some of these issues are described in the following sub-sections.



a) Flat plane b) Sphere

Figure 17: Scan of near-lambertian surfaces showing varying data density

5.5.1 Data acquisition and/or digitization

Data obtained from SL scanners is typically noisy and some data filtering must be performed to exclude data that does not represent the object. Software provided by the scanner manufacturers typically use proprietary methods to perform some level of filtering. Figure 17 shows a scan of a flat aluminum plate and a sphere, both of which have bead-blasted and diffused surfaces. It was observed that the data density over the surface varies. The data density is higher along the edges and at locations that have slightly higher reflectivity than the rest of the surface. Processing datasets that have non-uniform point density results in errors while calculating quality of parameters such as plane flatness, sphere diameter and form errors. It is also not clear whether these datasets were natively generated as a polygonal mesh or as 3D point cloud data.

5.5.2 Post-processing

Care must be taken while applying data filters in post-processing, as rejection of valid points or inclusion of invalid points will indirectly affect the performance results of the instrument. Structured light scanners produce varying point densities in the scanner work volume based on the location and orientation of the object surface. Two kinds of tests were performed that will show the effect of point cloud processing on scanner errors.

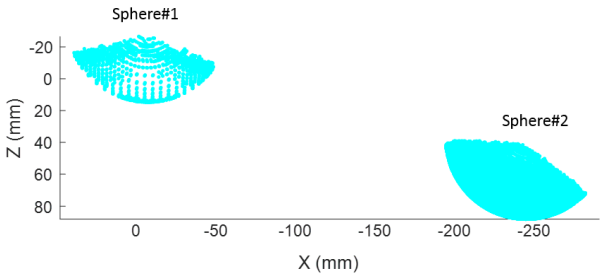


Figure 18: Scan data of two spheres at two different distances from the scanner

The first test was conducted by scanning two spheres, 15 times, at two distances from scanner SA as shown in Figure 18. Scan data of Sphere#1 had an average of 1465 points and Sphere#2 had an average of 13546 points. The average error in radius for Sphere#1 was 1.1 µm and that for Sphere#2 was 0.3 µm. Both the spheres were very similar in construction but resulted in different values for the error in radius. This was due to the varying point density and/or the location of the sphere in the scanner work volume.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

A second test was conducted by scanning a square flat plane of 150 mm width at 10 different angles, and five repeat scans were obtained at each angle. This data was fit to a plane and root-mean-square (RMS) value of the residuals was obtained and shown in Figure 19. This resulted in lower RMS noise of its residuals and the number of points scanned were lowered when the angle of rotation exceeded 15° to the scanner's XY plane. It should be noted here that the scanner's coordinate system was centered at the instrument work volume, the Z axis was away from the scanner in the horizontal plane and the XY plane is oriented vertically and not parallel to either the camera's image plane or the projector's sensor plane.

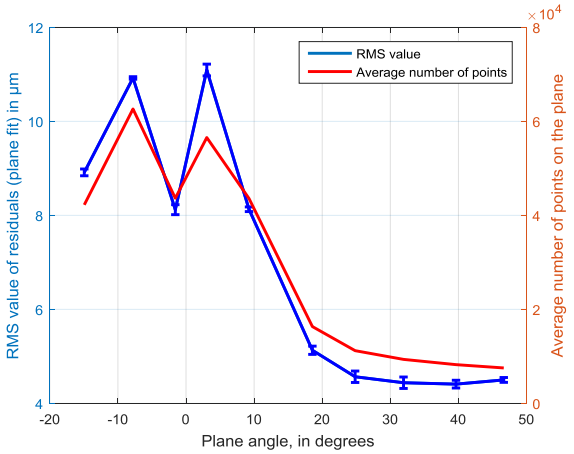


Figure 19: Noise and number of points on a flat plane at various angles

6 SUMMARY & FUTURE WORK

Structured light scanners are increasingly being used in many industrial applications. Several organizations are procuring these scanners based on limited understanding of their performance and a lack of adequate standards. Some guidance is provided by the VDI/VDE 2634 guidelines, but these guidelines need to be extended to encompass various kinds of SL scanners that are currently available.

DMG at NIST has started to explore these scanners by first understanding the sources of errors in the scanners and their effect on dimensional measurements. There are several error sources described in this paper that extend beyond the mathematical model of the scanner. The effect of each of these sources of errors on the quality parameters described in VDI/VDE 2634 guidelines is not well understood. Future activities are planned to identify the quality factors that are sensitive to these error sources and develop test procedures based on such information. The authors also plan to work with the ISO and other U.S. standards development organizations to develop new and fair performance evaluation tests that will improve commerce.

7 DISCLAIMER

Commercial equipment and materials may be identified to specify certain procedures. In no case does such identification imply recommendation or endorsement by the NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

8 ACKNOWLEDGEMENTS

The authors would like to thank the following experts and researchers for their valuable time, feedback and useful discussions: Dr. Craig Shakarji, Dr. Steve Phillips and Ms. Geraldine Cheok of NIST; Dr. Octavio Icasio Hernández of CENAM; Dr. Gabriel Taubin of Brown University; Mr. Luc Cournoyer of NRC-Canada; Mr. John Palmateer; Mr. Sean Woodward of NPL-UK; Dr. Michael McCarthy of Engineering Metrology Solutions – UK.

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

9 REFERENCES

[1] VDI-Standard: VDI/VDE 2634 Blatt 2: https://www.vdi.eu/guidelines/vdivde_2634_blatt_2-optische_3_d_messsysteme_bildgebende_systeme_mit_flaechenhafter_antastung/

[2] VDI-Standard: VDI/VDE 2634 Blatt 3: https://www.vdi.eu/guidelines/vdivde_2634_blatt_3-optische_3_d_messsysteme_bildgebende_systeme_mit_flaechenhafter_antastung/

[3] Taubin, G., Moreno, D., Lanman, D, "The Laser Slit 3D Scanner" http://mesh.brown.edu/desktop3dscan/ch4-slit.html (Accessed on 11/23/2018)

[4] El-Hakim S., Beraldin J., Blais, F., “Comparative evaluation of the performance of passive and active 3D vision systems”, Proceedings Volume 2646, Digital Photogrammetry and Remote Sensing, December 1995. https://doi.org/10.1117/12.227862

[5] MacKinnon, D., Beraldin, J., Cournoyer, J., Carrier, B., "Hierarchical characterization procedures for dimensional metrology," Proc.SPIE 7864, Three-Dimensional Imaging, Interaction, and Measurement, 786402 (27 January 2011); https://doi.org/10.1117/12.872124

[6] Beraldin, J., Carrier, B., MacKinnon D., Cournoyer, L., "Characterization of Triangulation-based 3D Imaging Systems using Certified Artifacts", NCSLI Measure, 7:4, 50-60, https://doi.org/10.1080/19315775.2012.11721620

[7] McCarthy, M., Brown, S., Evenden, A., & Robinson, A., "NPL freeform artefact for verification of non-contact measuring systems", Proc.SPIE 7864, Three-Dimensional Imaging, Interaction, and Measurement, 786402 (27 January 2011); http://doi.org/10.1117/12.876705

[8] Taubin, G., Moreno, D., Lanman, D, "3D Scanning with Structured Light" http://mesh.brown.edu/desktop3dscan/ch5-structured.html (Accessed on 11/23/2018)

[9] Zhang, Z., "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), Nov 2000. https://doi.org/10.1109/34.888718

[10] Bouget, J., "Camera Calibration Toolbox for MATLAB", http://www.vision.caltech.edu/bouguetj/calib_doc/ (Accessed 2/12/2019)

[11] Hartley, R., Zisserman, A., "Multiple View Geometry in Computer Vision", 2nd edition, Cambridge University Press New York, NY, USA, 2003 ISBN:0521540518

[12] Bradski, G., Kaehler, A., "Learning OpenCV: Computer Vision with the OpenCV Library", O'Reilly Media, Inc., 2013, ISBN:9780596516130 pp 373,376

[13] Falcao, G., Hurtos, N., Massich, J., Fofi, D., "Projector-Camera Calibration Toolbox", https://github.com/davidfofi/procamcalib.git, 2009.

[14] Moreno, D., and Taubin, G., "Simple, Accurate, and Robust Projector Camera Calibration”, In 3DIMPVT, pp. 464–471, 2012

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
”Sources of Errors in Structured Light 3D Scanners.”
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

[15] Huang, B., Ozdemir, S., Tang, Y., Liao, C., Ling, H."A Single-shot-per-pose Camera-Projector Calibration System for Imperfect Planar Targets", 2018 https://arxiv.org/pdf/1803.09058.pdf

[16] Iwai D., Mihara S., Sato K., "Extended Depth-of-Field Projector by Fast Focal Sweep Projection", IEEE Transactions on Visualization and Computer Graphics, 21 (4), Apr 18, 2015.

[17] Geng J., "Structured-light 3D surface imaging: a tutorial," Adv. Opt. Photon. 3, 128-160 (2011) https://doi.org/10.1364/AOP.3.000128

[18] Salvi, J., Pages, J., Batlle, J., "Pattern Codification Strategies in Structured Light Systems", Pattern Recognition, Vol 37, Issue 4, April 2004.

[19] Jeon, J., Choi, B., Kim, C., Kim, J., Kim, H., Kim, W., " Three-dimensional evaluation of the repeatability of scanned conventional impressions of prepared teeth generated with white- and blue-light scanners", The Journal of Prosthetic Dentistry, 114 (4), Oct 2015.

[20] Baker, M., Xi, J., Chicharo, J., Li, E., "A contrast between DLP and LCD digital projection technology for triangulation-based phase measuring optical profilometers", Proceedings Volume 6000, Two- and Three-Dimensional Methods for Inspection and Metrology III; 60000G (2005) https://doi.org/10.1117/12.632582

Rachakonda, Prem; Muralikrishnan, Balasubramanian; Sawyer, Daniel.
"Sources of Errors in Structured Light 3D Scanners."
Paper presented at SPIE Defense and Commercial Sensing 2019, Baltimore, MD, United States. April 14, 2019 - April 18, 2019.

A Workflow for Imaging 2D Materials using 4D STEM-in-SEM

Benjamin W. Caplins[1*], Ryan M. White[1], Jason D. Holm[1] and Robert R. Keller[1]

[1.] Material Measurement Lab, National Institute of Standards and Technology, Boulder, CO, USA
* Corresponding author: benjamin.caplins@nist.gov

The grain structure and orientation of polycrystalline two-dimensional (2D) materials modulates their thermal, mechanical, and electrical properties. Thus, to rationally control the properties of 2D materials, the as-synthesized grain structure of the 2D materials must be routinely and reliably measured. Although different methods can be used to spatially map the grain structure/orientation of 2D materials, electron diffraction methods are generally regarded as the 'ground truth' because they can resolve these characteristics on the nanometer length scale.

In this contribution, a four-dimensional (real-space and k-space) scanning transmission electron microscopy method for use in the scanning electron microscope (4D STEM-in-SEM) is described. This technique can map the structure/orientation of 2D materials on a wide range of length scales (10^{-2} m - 10^{-9} m). Briefly, a 30 keV focused electron beam is incident on the electron transparent 2D material. The electrons that transmit through the sample propagate in a field-free region and strike a phosphor screen forming a diffraction pattern. This (optical) diffraction pattern is then imaged onto a CCD camera. A home-built scan generator is used to raster the electron beam across the sample and trigger the storage of diffraction pattern images [1]. The low energy of the electron beam yields significant diffraction contrast while creating minimal knock-on damage [2].

Once collected, the output of a 4D STEM-in-SEM experiment can be regarded as a 'big' dataset which cannot be exhaustively inspected by humans – e.g., a single real-space scan (512 × 512) generates ~250000 diffraction patterns. We have developed a workflow based on a Fourier-space representation of the diffraction data to rapidly inspect 4D STEM-in-SEM data and extract information such as crystallographic orientation and identify distinct regions of the material (such as different thicknesses and interlayer rotations). A typical workflow for data inspection is as follows. First, since no de-scan coils exist in an SEM, the center of each diffraction pattern is located and de-scanned digitally. This step is particularly important for large fields-of-view. The diffraction pattern is then resampled to a polar grid, and the Fourier transform is applied to the angular coordinate. The transformed data can be visualized by creating colorized image of the amplitude and phase of the Fourier-space data. Since it can be reasonably assumed that the diffraction patterns of a 2D material are based on structures that have in-plane symmetry, we can initially restrict our inspection to just a few low-index Fourier components. These synthetic images allow a practical visualization of the 4D STEM-in-SEM dataset with minimal processing power and user time and can indicate what region/structures are present in the sample. Then a cross-correlation-based pattern matching approach can give higher quality orientation information. This Fourier representation may be a useful dimensionality reduction method when machine learning is used for further analysis.

Figure 1 shows examples of this Fourier based analysis applied to a graphene oxide sample. An image of the phase of a six-fold Fourier component is used to visualize the grain orientation, and the amplitude image to visualize multilayer structures. Figure 2 shows the technique applied to monolayer graphene. Inspection of the two-fold Fourier components unexpectedly reveals texture on the micrometer length

Caplins, Benjamin; White, Ryan; Holm, Jason; Keller, Robert.
"A Workflow for Imaging 2D Materials using 4D STEM-in-SEM."
Paper presented at Microscopy & Microanalysis 2019, Portland, OR, United States. August 4, 2019 - August 8, 2019.

scale that was assigned as polymer residue from a wet-transfer step. Furthermore, the analysis showed that the polymer was oriented with respect to the graphene lattice [3]. The workflow described here leverages the symmetry of 2D materials and the efficiency of the FFT and may be particularly helpful when the S/N and processing power is not available/necessary to perform a more detailed analysis [4].

References:

[1] W.C. Lenthe et al., Ultramicroscopy **195** (2018), p. 93.
[2] R.F. Egerton, Micron **119** (2019), p. 72.
[3] M. Gulde et al., Science **345** (2014), p. 200.
[4] Y. Han et al., Nano Letters **18** (2018), p. 3746.

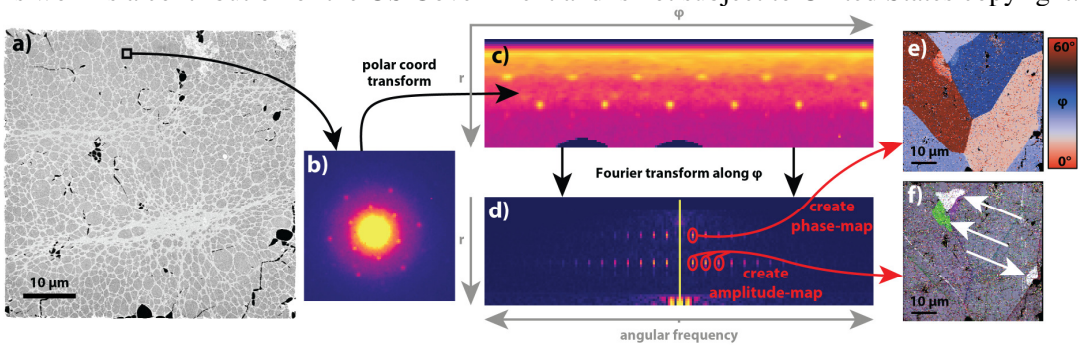


Figure 1. a) Secondary electron image of a mostly monolayer graphene oxide film on lacey carbon. b) A single diffraction pattern from a 4D STEM-in-SEM dataset. c) Diffraction pattern resampled to polar coordinates. d) Absolute value of the FFT (across φ) of the polar data. e) Colorized map of the phase of the circled Fourier component. f) Colorized map of the amplitudes of the circled Fourier components; distinct bilayer regions indicated with arrows.

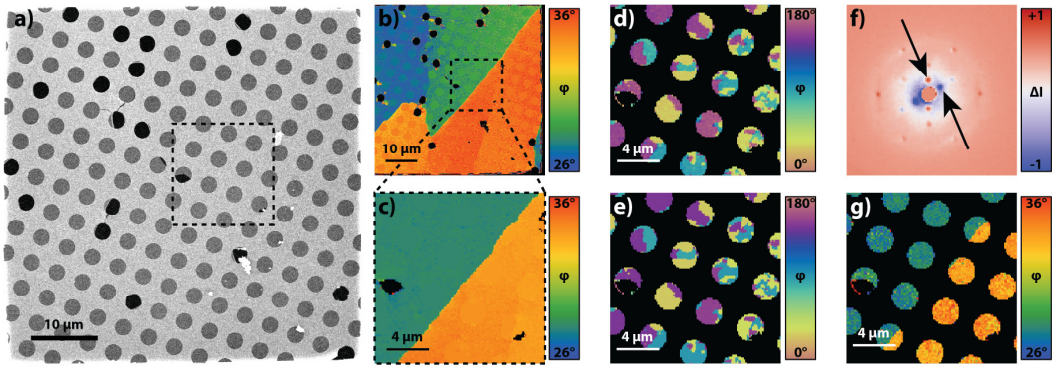


Figure 2. a) Secondary electron image of a supported monolayer graphene film. b) Colorized map of the phase of a six-fold Fourier component – i.e., graphene orientation. c) Zoomed-in orientation map derived using a cross-correlation approach. d) Colorized map of the phase of a two-fold Fourier component showing the texture of a polymer residue. e) Orientation map using a cross-correlation approach. f) The difference between the diffraction patterns in two polymer domains; the arrows indicate the polymer diffraction. g) The data in (e) *modulo* 60°, on the same colormap as (c), highlighting the orientation relationship.

Browser Fingerprinting using Combinatorial Sequence Testing

Bernhard Garn
SBA Research
Vienna, Austria
bgarn@sba-research.org

Dimitris E. Simos
SBA Research
Vienna, Austria
dsimos@sba-research.org

Stefan Zauner
FH Campus Wien
Vienna, Austria
stefan.zauner@stud.fh-campuswien.ac.at

Rick Kuhn
NIST
Gaithersburg, MD, USA
d.kuhn@nist.gov

Raghu Kacker
NIST
Gaithersburg, MD, USA
raghu.kacker@nist.gov

ABSTRACT

In this paper, we report on the applicability of combinatorial sequence testing methods to the problem of fingerprinting browsers based on their behavior during a TLS handshake. We created an appropriate abstract model of the TLS handshake protocol and used it to map browser behavior to a feature vector and use them to derive a distinguisher. Using combinatorial methods, we created test sets consisting of TLS server-side messages as sequences that are sent to the client as server responses during the TLS handshake. Further, we evaluate our approach with a case study showing that combinatorial properties have an impact on browsers' behavior.

KEYWORDS

combinatorial testing, security testing, browser fingerprinting

ACM Reference Format:
Bernhard Garn, Dimitris E. Simos, Stefan Zauner, Rick Kuhn, and Raghu Kacker. 2019. Browser Fingerprinting using Combinatorial Sequence Testing. In *Hot Topics in the Science of Security Symposium (HotSoS), April 1–3, 2019, Nashville, TN, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3314058.3314062

1 INTRODUCTION

The security and threat landscape of the computer systems of today can be viewed from a macroscopic and a microscopic point of view. Both views are essential, independent and influence each other. In this work, we focus on fingerprinting Internet browsers by analyzing their behavior when they are processing non-standard response messages from a server during a TLS handshake. These non-standard messages are derived using combinatorial methods and no other means than the resulting behavior are used in the fingerprinting approach. Browsers, a type of end-user software which is paramount, constitute the central piece of software by which computing power and the Internet are consumed on a variety of mobile and classic devices. A lot of effort is spent to annotate vulnerabilities found not only with the specific software where the vulnerability was discovered, but to also determine all other software that is also affected. In a *defensive* position, these efforts usually lead to security advisories where users are then asked to update the affected software to a version for which this vulnerability has been resolved. From an *offensive* or penetration testing position, a list of pairs consisting of a software-vulnerability combination provides the means for planning concrete penetration tests on systems. A prerequisite to this step is, however, that sufficient knowledge is available about the target system so that a detailed software version-vulnerabilities list can be used effectively. In such a hypothetical scenario, the process likely begins with a reconnaissance phase to determine the exact versions of the software that is running on the target system. To this end, in this paper, we employ a combinatorial sequence testing technique in a black-box testing approach tailored to fingerprint software in use. Specifically, we assume the role of a classical web server where users of that service want to connect with a secure connection (HTTP over TLS 1.2 [4]). Our idea is that by responding with certain TLS handshake messages (on some connection attempts), the resulting error messages (if any are sent) can be used to uniquely identify the used browser. After a successful fingerprinting of the browser, the next attack steps can be planned with the precise knowledge of the browser version in use on the target (i.e., exploits specifically developed for this target). We would like to note that it is possible to create a database of browser behavior depending on TLS sequences independent from any specific target and most importantly, in advance, and that it can be continuously updated to have it available for later usage.

Over the last couple of years, there has been a trend to offer more and more content over HTTPS, for a variety of reasons. Apart from the previously described use case, browser fingerprinting might be used in a variety of scenarios, like identifying users or obtaining information about the type and version of a browser in order to deliver suitable malware.

Our approach maps the behavior of browsers to *feature vectors*, upon which we base our classification. The resulting analysis on the obtained partitions of all considered browsers can be regarded as a way to *quantify* and *reason* about the *similarities* in observed browser behavior.

The goal of this work is *to investigate the applicability of combinatorial methods to generate artifacts upon which the behavior of browsers will be evaluated and recorded to be used as underlying means to build a fingerprinting approach based on it.* Specifically, the *research question* for this work reads as follows:

HotSoS, April 1–3, 2019, Nashville, TN, USA

ACM ISBN 978-1-4503-7147-6/19/04.
https://doi.org/10.1145/3314058.3314062

Garn, Bernhard; Simos, Dimitris; Zimmer, Stefan; Kuhn, David; Kacker, Raghu.
"Browser Fingerprinting using Combinatorial Sequence Testing."
Paper presented at Hot Topics in the Science of Security, Nashville, TN, United States. April 2, 2019 - April 3, 2019.

RQ: *Can the behavior observed from test sets created by sequence covering arrays containing TLS server-side messages be used to fingerprint browsers?*

The methods proposed in this paper can be used independently or in conjunction with existing methods and techniques for browser fingerprinting. Moreover, the research objective of this paper is not to increase the efficiency or to address specific limitations of existing approaches, but to evaluate the applicability of a new approach based on combinatorial methods to the problem of browser fingerprinting and as a result to extend the tools and techniques available in general in the field of fingerprinting.

We would like to note that some properties reported by a browser to a server (for example, its user agent string) are inherently not trustworthy (since they can be set easily arbitrarily), whereas the methods proposed in this paper is solely based on the observed behavior of browsers (i.e., the message-exchange occurring during a TLS handshake attempt). We expect it to be more difficult to defend against this type of fingerprinting or to fake the behavior on errors to imitate the error-behavior of another browser, as mitigation strategies would have to operate on the level of the underlying TLS implementation.

Contribution. In particular, this paper makes the following contributions:

- Proposes certain combinatorial sequences as test cases for fingerprinting browser behavior;
- Presents experimental results of a case study demonstrating our approach;

This paper is structured as follows. In Section 2 we discuss related work. We present our approach to fingerprinting of browsers using combinatorial methods in Section 3 and describe our developed testing framework in Section 4. In Section 5 we describe the setup of our case study and analyze the obtained results in Section 6. Last, Section 8 concludes the paper.

2 RELATED WORK

Approaches for browser fingerprinting often rely on properties exposed by the browser about itself and the underlying operating system. The authors of [5] followed this approach, and in [16] this methodology was strengthened by additionally considering browser plug-ins. In [1], the authors focused on fingerprinting scripts.

Another line of research uses not only properties, but also capabilities of browsers for developing fingerprinting approaches. A fingerprinting technique based on the onscreen dimension of font glyphs was presented in [7]. The rendering of 3D scenes together with text rendering in a web page on an HTML5 <canvas> element was used in [12] to base a fingerprinting method upon it. Other browser capabilities have also been used to create fingerprinting approaches [22].

In [13], the underlying JavaScript engine was used to devise a fingerprinting approach. The correlation between feature combinations and identification accuracy has been considered in [20].

In [17], the authors used planning with combinatorial methods for providing test cases for testing different TLS implementations.

3 COMBINATORIAL METHODS FOR FINGERPRINTING

In this section, we detail how combinatorial methods arising in the field of discrete mathematics in conjunction with an abstract modelling methodology can be used to create test sequences that enable fingerprinting of browsers. First, we describe the combinatorial structures employed in this paper, and then we present our modelling methodology and how the constructed sequences can be used for testing. While combinatorial methods have been used in the past in the context of combinatorial security testing [18], in this paper combinatorial methods are used for the first time as the underlying means to create a fingerprinting approach.

3.1 Sequence Covering Arrays

Sequence covering arrays (SCAs) [9],[2] are matrices designed to test software behavior that depends on the order of events, by ensuring that any t events will occur in every possible t-way order (allowing interleaving events among each subset of t events). A sequence covering array, $\mathrm{SCA}(N, S, t)$, is defined as an $N \times S$ matrix where entries are from a finite set S of s symbols, such that every t-way permutation of symbols from S occurs in at least one row and each row is a permutation of the s symbols [9]. The t symbols in the permutation are not required to be adjacent. That is, for every t-way arrangement of symbols $x_1, x_2, \ldots, x_t$, the regular expression $.*x_1.*x_2 \cdots .*x_t.*$ matches at least one row in the array.

For example, with six events, a, b, c, d, e, f, one subset of three events is $\{a, c, e\}$, which can be arranged in six possible permutations. There are[1] $\binom{6}{3} = 20$ sets of three events, and each can be arranged in $3! = 6$ orders. Using only 10 tests, it is possible to include all 3-way orders of these six events, as shown in Table 1. It can be shown that, for a given value of t, the number of tests required to cover all t-way orders grows with $\log S$. For the case where $t = 2, N = 2$ for all values of S, i.e., testing 2-way sequences never requires more than two tests, regardless of the number of events.

Sequence covering arrays have been used in a variety of testing applications, including industrial control systems [3], web applications [10], cryptographic hash functions [11], and laptop utilities [9]. They can be constructed using a simple greedy algorithm approach [9], although search and logic based strategies have been used as well [6],[8]. Greedy algorithms are generally faster, while other approaches may produce slightly smaller test arrays. To our knowledge, they have not been used for device fingerprinting or other applications outside of conventional software testing.

SCA generation. We implemented the algorithm given in [9] in the Python 2 programming language [15]. It takes as input the number of events n and the desired interaction strength t and returns a SCA for the specified configuration over the alphabet $\{0, \ldots, n-1\}$. The algorithm follows a greedy strategy, where after empty initialization of the test sequence set, in each iteration, a number of test sequences are created randomly, individually scored by the number of previously uncovered t-way sequences they cover, and then the highest scoring test sequence is added to the test set

[1] For two nonnegative integers n and k with $k \leq n$ we denote with $\binom{n}{k}$ the binomial coefficient. For a positive integer i we denote with $i!$ the factorial of i.

Garn, Bernhard; Simos, Dimitris; Zimmer, Stefan; Kuhn, David; Kacker, Raghu.
"Browser Fingerprinting using Combinatorial Sequence Testing."
Paper presented at Hot Topics in the Science of Security, Nashville, TN, United States. April 2, 2019 - April 3, 2019.

Table 1: All 3-event sequences of 6 events.

Test	Sequence
1	a b c d e f
2	f e d c b a
3	d e f a b c
4	c b a f e d
5	b f a d c e
6	e c d a f b
7	a e f c b d
8	d b c f e a
9	c e a d b f
10	f b d a e c

Table 2: Number of tests for generated SCAs.

n	t	#tests
2	2	2
3	2	2
3	3	6
4	2	2
4	3	8
4	4	24
5	2	2
5	3	10
5	4	28
5	5	120
6	2	2
6	3	10
6	4	36
6	5	156
6	6	720

until all required t-way sequences are covered. The sizes of the SCAs that this program generated are given in Table 2.

SCA verification. Test sequences used in this work were verified for coverage using the Combinatorial Sequence Coverage Measurement (CSCM) tool [23],[24]. CSCM was designed to allow testers to evaluate the sequence coverage of test sets that have not necessarily been generated to cover sequences. Event sequence testing has long been known to be important in fields such as communication protocols, and many methods exist to generate sequences that cover the state transition graphs for protocol testing. Many consumer-level applications, particularly for web sites and smartphone apps, also have the potential for complex interactions that are difficult to test fully. CSCM allows testers to determine the extent to which t-way sequences are covered in a test set, and to supplement existing tests as needed to enhance test thoroughness.

3.2 Application to fingerprinting

Now, we explain how to use SCAs in the context of browser fingerprinting. First, we briefly summarize some facts about the TLS protocol, which is used to establish a secure connection between two parties.

Transport Layer Security. The *transport layer security protocol* (TLS) can be used by browsers to establish a secure connection with a web server. The setup of this secure channel is achieved via the exchange of a sequence of messages, where the two parties negotiate some parameter values for later use. This procedure is encoded in the handshake protocol within the TLS specification, and both client and server should follow it. We regard all TLS messages that are specified for the server as a set of six abstract events $\mathcal{E}$, consisting of:

$$\mathcal{E} = \{\text{ServerHello}, \text{Certificate}, \text{ECDHEServerKeyExchange}, \quad (1a)$$
$$\text{ServerHelloDone}, \text{ChangeCipherSpec}, \text{Finished}\}. \quad (1b)$$

If we use the ordering derived from the message exchange in the TLS specification, then the following mapping of TLS events to numbers is canonical:

(0) ServerHello
(1) Certificate
(2) ECDHEServerKeyExchange
(3) ServerHelloDone
(4) ChangeCipherSpec
(5) Finished

We denote finite, nonempty sequence over the alphabet $\mathcal{E}$ is ascending order between angle brackets, for example $\langle 0, 3, 5\rangle$.

Combinatorial Sequence Model and derived test cases. Given a nonempty subset of cardinality κ of abstract TLS events[2] of the set $\mathcal{E}$, it is possible to construct a SCA for any strength $t \in \{1, \ldots, \kappa\}$. For later evaluation purposes, for any $\emptyset \neq E \subseteq \mathcal{E}$ of cardinality[3] κ we created and stored the image of E under the symmetric group of κ elements in the database. In other words, for any nonempty subset of sequences, we created all of its permutations and stored them in the database in the Sequence table and we denote the corresponding set of all test sequences with $\mathcal{S}$. It follows that we have in total

$$\sum_{i=1}^{6} \binom{6}{i} \cdot i! = 1956 \quad (2)$$

test sequences in the database.

Since we store, for any nonempty subset E of $\mathcal{E}$, all of its permutations in the database, it follows that we can find any SCA defined over the elements of E in the database (i.e., fetching exactly those rows from the database which correspond to the rows of the SCA).

For later use, we also fix an enumeration of the set $\mathcal{S}$, that is we fix a bijection[4] from the set $\{1, \ldots, |\mathcal{S}|\} \longrightarrow \mathcal{S}$.

Given a SCA, we refer to a sequence as a test sequence (i.e., row in the array) and to the array in its entirety as a test set. Such an abstract test sequence can now be translated into a concrete TLS message that will be sent to the client *over the network* by instantiating its values with default values taken from TLS attacker. We

[2]It would also be possible to consider not only subsets, but also sub-multisets, i.e., by allowing at least one event to appear strictly more than once in the considered selection. We leave this as future work.
[3]For a set S, we denote its cardinality with $|S|$.
[4]We obtained such an enumeration from the sequence identifiers in the database.

Garn, Bernhard; Simos, Dimitris; Zimmer, Stefan; Kuhn, David; Kacker, Raghu.
"Browser Fingerprinting using Combinatorial Sequence Testing."
Paper presented at Hot Topics in the Science of Security, Nashville, TN, United States. April 2, 2019 - April 3, 2019.

would like to emphasize that while we are considering permutations of subsets of the set $\mathcal{E}$, the concrete message values are taken from TLS attacker[5].

4 TESTING FRAMEWORK

In this section, we describe our testing setup of our automated framework. It is composed of several components and their general interactions are depicted in Figure 1. The frameworks' test execution component uses TLS attacker [19] for sending and receiving of TLS messages of clients (i.e., browsers). TLS test sequences are taken from a specified list, created using combinatorial methods. Subsequently, the logging component of the framework stores annotated results for each browser in a dedicated relational database. In other words, for any given browser, the framework executes the following steps:

(1) TLS attacker is started as a server and given as input the XML translation of a test sequence from a set test.
(2) The browser is instrumented to connect via HTTPS to a specific local url of TLS attacker.
(3) The framework (i.e., the server side in this connection attempt) will respond according to the encoded messages of the test sequence given in XML format.
(4) The exchange of TLS messages between the client (browser) and server will continue as long as possible until all messages from the test case have been sent from the server. The complete message exchange is recorded.
(5) The recorded message exchange is annotated and stored in a relational database.

Next, we describe each component in detail.

Test cases. In our experiments we considered all possible injective ordered sequences $\mathcal{S}$ over the alphabet $\mathcal{E}$ of length one to six. This means that, in particular, the set $\mathcal{S}$ has as a subset all t-SCAs for all $t \in \{1, 2, 3, 4, 5, 6\}$. In Section 6, during the evaluation, we will make use of this fact and compare the *distinguishing* capabilities of various subsets of $\mathcal{S}$, in particular those of SCAs for different event selections and different strengths. Due to the size advantage of SCAs compared to the set of all permutations of some fixed set of elements, it is clear that while for our initial experiments we were able to work the complete set $\mathcal{S}$, in future extension steps of an existing analysis database it would be more desirable to follow a test set augmentation strategy based on higher strength of SCAs instead of adding all permutations.

Test execution. The test execution component uses TLS attacker[6] [19], which is an open source framework that allows the creation of custom TLS message flows, both from a client and server side. Since we want to fingerprint clients (i.e., browsers), we use TLS attacker in the server role. For each abstract test sequence, an XML encoded TLS handshake sequence is created, which is one of the input formats of TLS attacker. If available, we start the browsers in headless mode. The browsers connect to a locally running TLS attacker. A self-signed root certificate was created and added to the list of trusted certificates for each browser and the certificate that is sent by TLS attacker was signed with the private key of the root certificate (no intermediate certificate). Also, the certificate sent by our server has a V3-Extension called `subjectaltname`. This was necessary since Google Chrome uses this extension to validate the certificate and otherwise it would not accept it.

During the handshake, the framework will reply with exactly the messages in the order specified in the current test sequence with the concrete message values, except for the certificate, instantiated to default values provided by TLS attacker[7]. The ciphersuite was fixed to `TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA`, as all tested browsers considered it secure.

The browser process is terminated as soon as the stream (from TLS attacker) is closed and all the results are stored. This is the same process for both headless and not headless.

Logging and database. While instrumenting TLS attacker, we parse its comprehensive output. This output includes detailed information about sent and received TLS messages. It also outputs TLS `alert messages` and prints out exceptions that happened during execution (stored in the field `ExceptionHandshake`), for example if the browser closed the socket or if TLS attacker failed to parse a received message. The tests also send an HTTPS-response to the browser, so it is possible to record exceptions that happened after the handshake (stored in the field `ExceptionPostHandshake`), when data should be sent over the now secure channel. We added this information to make it possible to have more data in the feature vectors used as distinguishers (e.g., no exception is thrown during the handshake, but the socket is closed as soon as TLS attacker tries to send data).

This logging information, together with the tested client (i.e., browser) is then stored in a SQLite database [14]. Moreover, the database also contains tables with

- the list of browsers, their paths and command line options for how to start them, respectively;
- the set of TLS events $\mathcal{E}$;
- a workflow skeleton that is populated with the saved TLS messages;
- the list of all considered test sequences $\mathcal{S}$.

We give some examples for results obtained in Table 3, as they are stored in the database.

Walk through example. We illustrate our approach and the test case execution with an example. Consider a five event selection out of the set $\mathcal{E}$ where the event Certificate is omitted, i.e. we choose the following sequence

$$\langle \text{ServerHello}, \text{Finished}, \text{ServerHelloDone}, \quad (3a)$$
$$\text{ChangeCipherSpec}, \text{ECDHEServerKeyExchange}\rangle \quad (3b)$$
$$= \langle 0, 5, 3, 4, 2\rangle \quad (3c)$$

built from these events. The corresponding XML encoded sequence is depicted in Listing 1.

```
1 <?xml version="1.0" encoding="UTF-8" standalone="
      yes"?>
2 <workflowTrace>
3     <Receive>
4         <expectedMessages>
```

[5] We leave their additional manipulation as future work.
[6] Release v2.6.
[7] https://github.com/RUB-NDS/TLS-Attacker/blob/master/TLS-Core/src/main/resources/default_config.xml

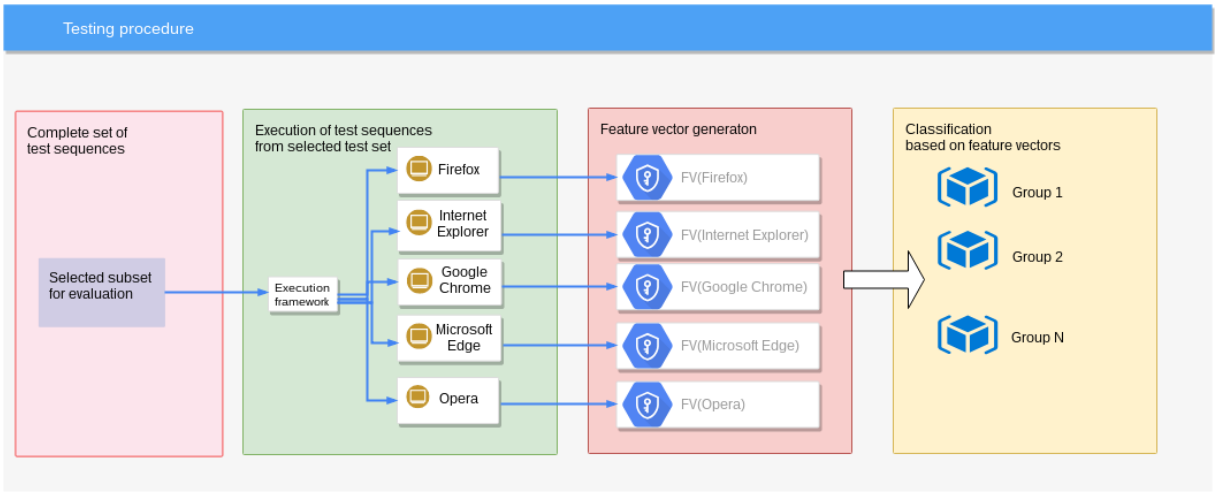


Figure 1: Overview of the fingerprinting process.

Table 3: Excerpt from the `Results` table.

ID	ReceivedMsgs1	ReceivedMsgs2	AlertMessage	ExceptionHandshake	ExceptionPostHandshake	br_id	Seq_id
6	CLIENT_HELLO	AlertMessage	UNEXPECTED_MESSAGE			1	2
761	CLIENT_HELLO	AlertMessage	UNEXPECTED_MESSAGE		java.net.SocketException: Connection reset by peer: socket write error	1	153
3881	CLIENT_HELLO			java.net.SocketException: Connection reset by peer: socket write error	java.net.SocketException: Connection reset by peer: socket write error	1	777
3882	CLIENT_HELLO				java.net.SocketException: Software caused connection abort: socket write error	2	777
3883	CLIENT_HELLO	AlertMessage	UNEXPECTED_MESSAGE		java.net.SocketException: Software caused connection abort: socket write error	3	777
3884	CLIENT_HELLO				java.net.SocketException: Software caused connection abort: socket write error	4	777
3885	CLIENT_HELLO	AlertMessage	UNEXPECTED_MESSAGE		java.net.SocketException: Software caused connection abort: socket write error	5	777

```
5            <ClientHello/>
6        </expectedMessages>
7    </Receive>
8    <Send>
9        <messages>
10           <ServerHello/>
11           <Finished/>
12           <ServerHelloDone/>
13           <ChangeCipherSpec/>
14       </messages>
15   </Send>
16   <Receive>
17       <expectedMessages>
18           <ECDHClientKeyExchange/>
19           <ChangeCipherSpec/>
20       <Finished/>
21   </expectedMessages>
22   </Receive>
23   <Send>
24       <messages>
25           <ECDHEServerKeyExchange/>
26       </messages>
27   </Send>
28   <Send>
29   <messages>
30           <HttpsResponse>
31       </HttpsResponse>
32   </messages>
33   </Send>
34 </workflowTrace>
```

Listing 1: TLS 1.2 altered handshake workflow.

For this example, we consider the case of testing the behavior of the browser Firefox. It is started in `headless` mode pointing to a local resource from the execution framework using the following command:

```
C:\Program Files\Mozilla Firefox\firefox.exe
        -headless -url https://localhost:5555
```

Upon execution, this test sequence led to the following exchange of TLS messages:

(1) `ClientHello` sent from Firefox.
(2) The framework sends the following messages:

⟨ServerHello, Finished, ServerHelloDone, (4a)
ChangeCipherSpec⟩ (4b)

(3) The client (Firefox in this case) responds and the response is identified by the execution framework as a TLS alert message of type `UNEXPECTED_MESSAGE`. Furthermore, we obtain

from TLS attacker an exception of type `java.net.SocketException: Connection reset by peer: socket write error`.
(4) Finally, an annotated version of the above test execution result is stored in the database.

5 CASE STUDY

In this section, we describe in detail the sequences used as test sets and the tested browsers and their version.

The objective of this case study is to provide initial experimental results about the capabilities of the executed methods for fingerprinting browsers. To this end, we ran the tests on a prevalent operating system for major browser vendors.

Test sets. As already mentioned in Section 3.2, our automated approach made it possible to work with the complete set $\mathcal{S}$ containing all permutations for all nonempty subset selections of TLS messages.

Independently, we used the generated SCAs (see Section 3.1) to obtain for every compatible selection of events a list of test sequence IDs. These IDs were used to obtain a subset of rows from the `Sequence` table in the database corresponding to the instantiated abstract SCA with the TLS messages selection.

Since we can find these SCAs as subsets of $\mathcal{S}$, by executing all test sequences of $\mathcal{S}$ against all browsers, we have also tested all SCAs of interest.

Browsers. In total, we tested five browsers for a case study, consisting of the following:

(1) Mozilla Firefox, version 64.0.0.6914;
(2) Opera, version 57.0.3098.106;
(3) Google Chrome, version 71.0.3578.98;
(4) Microsoft Internet Explorer, version 11.0.17134.1;
(5) Microsoft Edge, version 11.00.17134.471.

Framework setup. All experiments were performed on Windows 10 Pro, 64-bit Build 17134.472 Version 1803 running inside a virtual machine created with VMware Workstation 12 Pro 12.5.9 build-7535481. During the testing, all browsers connected to a locally running instance of TLS attacker.

6 EVALUATION

In this section, we present our results from running the experiments described in Section 5 and the theoretical criteria upon which we base our analysis. We explain how we instantiate feature vectors for abstract analysis in Section 6.1 and subsequently remark on how we compare them in Section 6.2. Afterwards, we elaborate on the results obtained in Section 6.3. The evaluation was performed with a program written in Perl v5.24.1 [21], which queried the results database and carried out the necessary steps for the analysis of our results.

Results show that the methods presented in this paper are effective for distinguishing between browser classes. That is, a very large number of the SCA tests were able to determine the browser type as belonging to one of the three categories: {Firefox}, {Google Chrome, Opera}, {Microsoft Internet Explorer, Microsoft Edge}.

6.1 Feature vector definition

For given browser and nonempty set of test sequences $S \subseteq \mathcal{S}$, we define a *feature vector* as follows:

- For each $s \in S$, let r_s be the result of executing test sequence s against the given browser (queried from the results database), stored as an array of strings. Note that the length of this array is uniform for all browsers and all test sequences in the set $\mathcal{S}$.
- Let fv denote an array of length $|S|$, where each entry is a reference to the array r_s, in ascending order according to the chosen enumeration of the set $\mathcal{S}$. The result can be interpreted as a two-dimensional array where the the first position of a two-dimensional index pair corresponds to the enumeration identifier of a test sequence and the second position to a column in the `Result` table schema.
- We define the array fv as feature vector for the given browser and set S of sequences.

We exemplify our definition of feature vectors with an example. Consider the browser Mozilla Firefox together with the following set of $S \subseteq \mathcal{S}$ of test sequences:

$$S = \{\langle \text{Certificate} \rangle, \langle \text{Finished}, \text{ChangeCipherSpec}, \text{ServerHello} \rangle\} \tag{5}$$

The result of these two test sequences executed against Firefox (which has `browser_id` equal to one) are depicted in Table 3, where

- the sequence ⟨Certificate⟩ has `Sequence_id` equal to two and the corresponding result is stored in the `Result` table with `ID` equal to six.
- the sequence ⟨Finished, ChangeCipherSpec, ServerHello⟩ has `Sequence_id` equal to 153 and the corresponding result is stored in the `Result` table with `ID` equal to 761.

The resulting feature vector fv has length two. The first entry points to an array of strings with the respective values shown in Table 3 for `ID` equal to six[8]:

```
("CLIENT_HELLO"|"AlertMessage"|"UNEXPECTED_MESSAGE"|""|"")
```
(6)

The second entry points to an array of strings with the respective values shown in Table 3 for `ID` equal to 761[9]:

```
("CLIENT_HELLO"|"AlertMessage"|"UNEXPECTED_MESSAGE"|""|  (7a)
                              "java.net.SocketException:  (7b)
          Connection reset by peer: socket write error") (7c)
```

6.2 Classification of feature vectors

Since we are interested in finding nonempty subsets of the set $\mathcal{S}$ where we can observe *different behavior* of the tested browsers, we now give a precise definition of how the term *different behavior* is to be understood in this paper. Suppose we are given a nonempty subset $S \subseteq \mathcal{S}$ and two different browsers, then we say that they

[8] The characters (and) denote the start and end of the array, respectively; the double high quotes delimit strings; and the character | is used as array element separator.
[9] See Footnote 8.

exhibit different behavior with respect to S, if and only if, their respective feature vectors differ[10].

For five browsers, there are ten possible pairwise comparisons between their behaviors (i.e., pairwise comparisons between their respective feature vectors). We use these pairwise comparisons to define an equivalence relation on the set of all browsers. For fixed nonempty set $S \subseteq \mathcal{S}$, two browsers are equivalent, if and only if, they exhibit the same behavior. In other words, two browsers are members of the same equivalence class, if and only if, they exhibit the same behavior for the set S. Due to our interest in fingerprinting browsers according to their behavior using test cases created with combinatorial methods, we are especially interested in analyzing the resulting partitions for different nonempty sets of test sequences.

6.3 Analysis of results

Firstly, we make some general observations on our results in Section 6.3.1. Then, we proceed with our analysis for groups of sequences of the same length for lengths from 1 up to 6 in Section 6.3.2 until Section 6.3.7. For $i \in \{1, 2, 3, 4, 5, 6\}$ we denote with S_i the subset of $\mathcal{S}$ consisting of exactly those sequences of length i. Note that in the case of $n = t$, the notions of the image of a set of cardinality n under the full permutation group and a *SCA* of strength n result in the same set of sequences.

6.3.1 General observations. For each individual test sequence in the set $\mathcal{S}$, we have seen at least the following pairwise equalities between behavior of two specific selections of two browsers:

- The browsers Microsoft Internet Explorer and Microsoft Edge exhibit the same behavior.
- The browsers Google Chrome and Opera exhibit the same behavior.

It follows that:

- Both of these selections of two browsers will have the same behavior for any nonempty subset of the set $\mathcal{S}$.
- For both of these selections of two browsers, for any nonempty set of test sequences, the resulting partition will have at most three classes.
- The approach for fingerprinting presented in this paper is currently not able to distinguish browsers within these two browser pairs.
- The result that those two browser-pair selections always exhibit the same behavior is not surprising, since they internally use closely related libraries for handling TLS handshakes.

It is possible to make the above statements more precise, which we state in the form of an explicit description of the appearing partitions. For each test sequence, the number of equivalence classes in the corresponding partition is an element of the set

$$C = \{1, 2, 3\}, \tag{8}$$

and for each number in the set C there is at least one test sequence where the partition corresponding to this sequence (i.e., singleton of test sequence selection) has exactly this number of equivalence classes.

For each number in the set C, we give now a more detailed description for the occurring partitions.

- Number of equivalence classes equal to one: All browsers have the same behavior and the corresponding partitions are equal. This partition occurs seven times.
- Number of equivalence classes equal to two: There are two different partitions:
  - One partition consisting of the two classes:
    (1) {Firefox, Google Chrome, Opera},
    (2) {Microsoft Internet Explorer, Microsoft Edge}
    occurring 22 times.
  - The other partition consisting of the two classes:
    (1) {Firefox},
    (2) {Microsoft Internet Explorer, Microsoft Edge, Google Chrome, Opera},
    occurring only once.
- Number of equivalence classes equal to three: We denote this unique class as $\mathcal{P}_3$ and the classes for the respective partitions are equal to:
  (1) {Firefox},
  (2) {Google Chrome, Opera},
  (3) {Microsoft Internet Explorer, Microsoft Edge}.
  This case occurs 1926 times.

After this analysis of all possible singletons of test sequence selections, next we analyze test sets for the same length and of cardinality at least two, in particular SCAs for different strengths.

6.3.2 Sequences of length 1. There are six selections of one event.

⟨*ServerHello*⟩*:* In this case, the resulting partition contains only one class with five elements, i.e., all browsers behave the same way.

⟨*Certificate*⟩, ⟨*ECDHEServerKeyExchange*⟩, ⟨*ServerHelloDone*⟩, ⟨*ChangeCipherSpec*⟩, ⟨*Finished*⟩*:* All of these cases resulted in the same partition of the set of all browsers, which has the following structure:

- Class I: {Microsoft Internet Explorer, Microsoft Edge}
- Class II: {Firefox, Google Chrome, Opera}

In the case of singleton selections of test sequences of length one, the concepts of test sequence, SCA of strength one and image under the full permutation group all coincide. We conclude that different singleton selections (i.e., different selections of one event), have different differentiation capabilities.

6.3.3 Sequences of length 2. For all subset selections of cardinality two, for our generated SCAs of strength two, the result is the partition $\mathcal{P}_3$.

6.3.4 Sequences of length 3. For all subset selections of cardinality three, for all our generated SCAs of strengths $t \in \{2, 3\}$, the result is the partition $\mathcal{P}_3$.

6.3.5 Sequences of length 4. For all subset selections of cardinality four, for all our generated SCAs of strength $t \in \{2, 3, 4\}$, the result is the partition $\mathcal{P}_3$.

[10] Equality of feature vectors is to be understood as canonical equality between two-dimensional arrays with the same dimensions; i.e., in each position equality for the respective strings holds.

Garn, Bernhard; Simos, Dimitris; Zimmer, Stefan; Kuhn, David; Kacker, Raghu.
"Browser Fingerprinting using Combinatorial Sequence Testing."
Paper presented at Hot Topics in the Science of Security, Nashville, TN, United States. April 2, 2019 - April 3, 2019.

6.3.6 Sequences of length 5. For all subset selections of cardinality five, for all our generated SCAs of strength $t \in \{2, 3, 4, 5\}$, the result is the partition $\mathcal{P}_3$.

6.3.7 Sequences of length 6. For all subset selections of cardinality six, for all our generated SCAs of strength $t \in \{2, 3, 4, 5, 6\}$, the result is the partition $\mathcal{P}_3$.

6.4 Interpretation of results

An analysis of the results in the previous evaluation shows that testing with the selection of only one test sequence of only one event leads to the weakest results in terms of differentiation. It is interesting to note that for some selections of only individual test sequences consisting of at least two events, the resulting partition is equal to $\mathcal{P}_3$. This case even occurs for about 98% of all individual event sequence selections. When considering test sequence selections of cardinality at least two, the resulting partition is also always equal to $\mathcal{P}_3$ and the best distinguishing capabilities obtained in this paper are reached. In particular, whenever there are at least two events appearing in the test sequence and a SCA is chosen as test set, then the best possible partition $\mathcal{P}_3$ is obtained.

Based on these results, we can answer the *research question* regarding the applicability of combinatorial methods to the problem of fingerprinting browsers in the affirmative.

7 THREATS TO VALIDITY

In this section, we comment on possible threats to validity of this work. It is clear that the performed case study is limited, since it only contains five browsers all running on Windows 10 Pro. Another threat stems from the fact that we relied on TLS attacker to act as the server side when the clients (i.e., browsers) attempted the TLS handshake. Likewise, although the goal of the paper is to investigate the applicability of combinatorial methods for fingerprinting, it is clear that a comparison of the proposed methods in this paper with existing approaches for browser fingerprinting would help to properly classify and position this work into the existing literature and methodologies for browser fingerprinting. We plan to conduct such a comparison in future work.

8 CONCLUSION AND FUTURE WORK

In this paper, we used combinatorial methods to create finite non-empty sequences of TLS messages to be used as the underlying means for a method of browser fingerprinting. An experimental evaluation shows that test sets of TLS sequences where the defined order of events compared to the specification is changed, lead to differentiation capabilities.

However, our results also showcase that a refined modelling is needed to strengthen the approach. This model development could be done in parallel to the extension of our set of tested browsers. As briefly mentioned before, the additional manipulation of TLS message contents and the extension to also consider multi-sets of events are directions for future research. Finally, any difference in observed behavior could be analyzed from the point of view of conformance testing, linking browser fingerprinting to problems in the field of conformance testing like undefined behavior or conformance quantification.

ACKNOWLEDGMENTS

The research presented in this paper was carried out in the context of the Austrian COMET K1 program and partly publicly funded by the Austrian Research Promotion Agency (FFG) and the Vienna Business Agency (WAW).

Moreover, this work was performed partly under the following financial assistance award 70NANB18H207 from U.S. Department of Commerce, National Institute of Standards and Technology.

Disclaimer: *Products may be identified in this document, but identification does not imply recommendation or endorsement by NIST, nor that the products identified are necessarily the best available for the purpose.*

REFERENCES

[1] Gunes Acar, Marc Juarez, Nick Nikiforakis, Claudia Diaz, Seda Gürses, Frank Piessens, and Bart Preneel. 2013. FPDetective: Dusting the Web for Fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*. ACM, New York, NY, USA, 1129–1140. DOI: http://dx.doi.org/10.1145/2508859.2516674

[2] Y. Chee, C. Colbourn, D. Horsley, and J. Zhou. 2013. Sequence Covering Arrays. *SIAM Journal on Discrete Mathematics* 27, 4 (2013), 1844–1861. DOI: http://dx.doi.org/10.1137/120894099 arXiv:https://doi.org/10.1137/120894099

[3] G. Dhadyalla, N. Kumari, and T. Snell. 2014. Combinatorial Testing for an Automotive Hybrid Electric Vehicle Control System: A Case Study. In *2014 IEEE Seventh International Conference on Software Testing, Verification and Validation Workshops*. 51–57. DOI: http://dx.doi.org/10.1109/ICSTW.2014.6

[4] Tim Dierks and Eric Rescorla. 2008. RFC 5246 - The Transport Layer Security (TLS) Protocol Version 1.2. https://tools.ietf.org/html/rfc5246. (2008). Accessed: 2019-01-07.

[5] Peter Eckersley. 2010. How Unique Is Your Web Browser?. In *Privacy Enhancing Technologies*, Mikhail J. Atallah and Nicholas J. Hopper (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–18.

[6] Esra Erdem, Katsumi Inoue, Johannes Oetsch, Jörg Pührer, Hans Tompits, and Cemal Yılmaz. 2011. Answer-set programming as a new approach to event-sequence testing. (2011).

[7] David Fifield and Serge Egelman. 2015. Fingerprinting Web Users Through Font Metrics. In *Financial Cryptography and Data Security*, Rainer Böhme and Tatsuaki Okamoto (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 107–124.

[8] M. Z. Mohd Hazli, Z. Kamal Z., and O. Rozmie R. 2012. Sequence-based interaction testing implementation using Bees Algorithm. In *2012 IEEE Symposium on Computers Informatics (ISCI)*. 81–85. DOI: http://dx.doi.org/10.1109/ISCI.2012.6222671

[9] D. R. Kuhn, J. M. Higdon, J. F. Lawrence, R. N. Kacker, and Y. Lei. 2012. Combinatorial Methods for Event Sequence Testing. In *2012 IEEE Fifth International Conference on Software Testing, Verification and Validation*. 601–609. DOI: http://dx.doi.org/10.1109/ICST.2012.147

[10] H. Mercan and C. Yilmaz. 2014. Pinpointing Failure Inducing Event Orderings. In *2014 IEEE International Symposium on Software Reliability Engineering Workshops*. 232–237. DOI: http://dx.doi.org/10.1109/ISSREW.2014.34

[11] N. Mouha, M. S. Raunak, D. R. Kuhn, and R. Kacker. 2018. Finding Bugs in Cryptographic Hash Function Implementations. *IEEE Transactions on Reliability* 67, 3 (Sep. 2018), 870–884. DOI: http://dx.doi.org/10.1109/TR.2018.2847247

[12] Keaton Mowery and Hovav Shacham. 2012. Pixel perfect: Fingerprinting canvas in HTML5. *Proceedings of W2SP* (2012), 1–12.

[13] Martin Mulazzani, Philipp Reschl, Markus Huber, Manuel Leithner, Sebastian Schrittwieser, Edgar Weippl, and FC Wien. 2013. Fast and reliable browser identification with javascript engine fingerprinting. In *Web 2.0 Workshop on Security and Privacy (W2SP)*, Vol. 5. Citeseer.

[14] SQLite project. 2019. SQLite. https://www.sqlite.org/index.html. (2019). Accessed: 2019-01-07.

[15] Python Software Foundation. 2019. Python. https://www.python.org/. (2019). Accessed: 2019-01-07.

[16] Research project of the Electronic Frontier Foundation. 2019. Panopticlick. https://panopticlick.eff.org/. (2019). Accessed: 2019-01-07.

[17] Dimitris E Simos, Josip Bozic, Bernhard Garn, Manuel Leithner, Feng Duan, Kristoffer Kleine, Yu Lei, and Franz Wotawa. 2018. Testing TLS using planning-based combinatorial methods and execution framework. *Software Quality Journal* (2018), 1–27.

[18] D. E. Simos, R. Kuhn, A. G. Voyiatzis, and R. Kacker. 2016. Combinatorial methods in security testing. *IEEE Computer* 49 (2016), 40–43.

[19] Juraj Somorovsky. 2016. Systematic Fuzzing and Testing of TLS Libraries. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, New York, NY, USA, 1492–1504.

[20] Kazuhisa Tanabe, Ryohei Hosoya, and Takamichi Saito. 2019. Combining Features in Browser Fingerprinting. In *Advances on Broadband and Wireless Computing, Communication and Applications*, Leonard Barolli, Fang-Yie Leu, Tomoya Enokido, and Hsing-Chung Chen (Eds.). Springer International Publishing, Cham, 671–681.
[21] The Perl Foundation. 2019. Perl. https://www.perl.org/. (2019). Accessed: 2019-01-07.
[22] Thomas Unger, Martin Mulazzani, Dominik Fruhwirt, Markus Huber, Sebastian Schrittwieser, and Edgar Weippl. 2013. Shpf: Enhancing http (s) session security with browser fingerprinting. In *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on*. IEEE, 255–261.
[23] Z. B. Ratliff. 2018. Black-box Testing Mobile Applications Using Sequence Covering Arrays. (2018). undergraduate thesis, Texas A&M University.
[24] Zachary Ratliff. 2019. CSCM-Tool. https://github.com/zachratliff22/CSCM-Tool. (2019). Accessed: 2019-01-07.

Garn, Bernhard; Simos, Dimitris; Zimmer, Stefan; Kuhn, David; Kacker, Raghu.
"Browser Fingerprinting using Combinatorial Sequence Testing."
Paper presented at Hot Topics in the Science of Security, Nashville, TN, United States. April 2, 2019 - April 3, 2019.

Applications of machine learning at the limits of form-dependent scattering for defect metrology

Mark-Alexander Henn*, Hui Zhou, Richard M. Silver, and Bryan M. Barnes

Nanoscale Device Characterization Division, Physical Measurement Laboratory,
National Institute of Standards and Technology, 100 Bureau Drive MS 8212,
Gaithersburg, MD, USA 20899-8212

ABSTRACT

Undetected patterning defects on semiconductor wafers can have severe consequences, both financially and technologically. Industry is challenged to find reliable and easy-to-implement methods for defect detection. In this paper we present robust machine learning techniques that can be applied to classify defect images. We demonstrate the basic principles of an algorithm that uses a convolutional neural network and discuss how such networks can be improved not only in their architecture but also tailored to the specific challenges of defect inspection through more specialized performance metrics. These advances may lead to more cost-efficient measurements by adjusting the decision threshold to minimize the number of wrong defect detections.

Keywords: electromagnetic simulation, sensitivity and uncertainty evaluation, through-focus three-dimensional field

1. INTRODUCTION

The challenges of optical defect inspection with decreasing feature dimensions, increasing layout complexity, and increasing materials variation have led recently to explorations of viable wavelengths shorter than 193 nm, a deep ultraviolet (DUV) wavelength. Experimental prototyping and calculations have been reported in the vacuum ultraviolet (VUV) [1]. Additional simulations have been perfomed by our group for wavelengths in the DUV, VUV, and extreme ultraviolet (EUV)[2, 3]. Note that the optical scattering from defects in layouts patterned using uniaxial lithography is enhanced due to form birefringence [4] if the pitch $p \gg \lambda$, the inspection wavelength. Whenever wavelengths are considered where $p \leq \lambda$, such simulations yield form-dependent high-spatial frequency scattering; a forthcoming manuscript will explore the implications and limits of this form-dependent scattering.

The semiconductor metrology literature is just beginning to reflect the growing need for data driven approaches such as machine learning (ML). Examples thus far include the measurement of critical dimensions [5], and inspection of defects in through-silicon vias [6], EUV masks [7], and patterned defects [8] as well as wafer mapping of patterned defects [9]. Growth in these areas should not be surprising, given the prominence of ML based research for process control in such diverse fields as wood [10], fabric inspection [11], hard-disk drive manufacturing [12, 13], and more general image recognition tasks [14–16]. Given the severe impacts, both financial and technological, that undetected defects on wafers can cause, it is worth investigating how the defect metrologist's toolbox can be extended using ML techniques and how the challenge of patterned defect inspection may differ from more general pattern recognition tasks.

In this paper, using the results of a simulation study that employs realistic wafer design, we present how an algorithm that is using a convolutional neural network can be a reliable tool for defect detection. Panels (a-d) of Fig. 1 show a schematic representation of the simulation domain for a defect-free and defect-containing layout, respectively. The layout is based on publically available information about recent manufacturing processes [17]. For simplicity, the details of the simulation geometry including materials constants and geometry files to reproduce the results here have been deposited elsewhere [18]. In the following we use images based on electromagnetic simulations for an inspection wavelength of $\lambda = 157$ nm using our in-house implementation [19, 20] of a finite-difference time-domain [21] (FDTD) model. In this work, we assume a normal angle of incidence and a polarization that is orthogonal to the direction of the defect, i.e., polarization along the x-axis in Fig. 1. To account for the measurement (shot) noise, we apply Poisson noise [22] to the raw images based on a photon density of 10 nm^{-2}.

*mark.henn@nist.gov

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

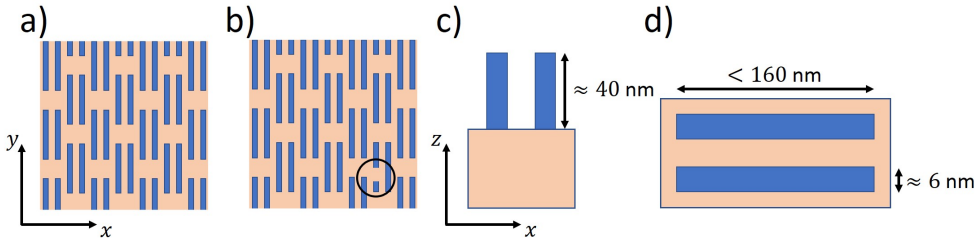


Figure 1. Schematic representation of a) ideal layout, b) defect-containing layout with an intrusion, and c-d) dimensions of unit cell.

In contrast to papers previously published by our group that based the defect detection on intensity thresholding [23], mean difference intensity [24], signal-to-noise ratio (SNR) [2, 3] or linear classification algorithms [8] for differential images or even differential volumes [25–27], we use the wavelet-based compressed raw images as an input to our algorithms. We use the compressed images in order to reduce the data size and the corresponding memory requirements without losing important details of the images. Stated differently, our prior publications made heavy use of defect-containing and defect-free images that could be subtracted from one another, but in this work we do not utilize difference images or intensity histograms thus more realistically addressing this challenging problem. Here, we present the basic principles of convolutional neural networks, demonstrating methodologies to improve upon an example from the literature [28] to illustrate specifically the several algorithmic design considerations available - and required - for applying ML to defect metrology.

2. DEFECT DETECTION USING CONVOLUTIONAL NEURAL NETWORKS

2.1 Basics of Machine Learning

Many defect detection algorithms can be understood as optimization problems, in which we start with a set of labeled data of size N, in our case images $x^{(i)} \in \mathbb{R}^{27\times 24}$ and corresponding labels $y^{(i)} \in \{0,1\}$, $i = 1,\ldots,N$ that are either 0 (if the image shows a defect) or 1 (if it does not). Based on this set we want to construct a multi-parameter prediction function, $f(\theta, \{y^{(i)}\}, \{x^{(i)}\})$ that shall be used to classify images that do not belong to the original set of labeled data. The values of the parameters θ of this prediction function are determined by optimization. Based on an adequate function $\mathcal{L}$ that can be used to measure the quality of the prediction, usually a function measuring the difference between the ground-truth label and the predicted label, also called a loss-function, we seek to minimize

$$\mathcal{L}\left[y^{(j)} - f(\theta, \{y^{(i')}\}, \{x^{(i')}\})\right],\ j \in \mathcal{A},\ i' \in \{1,2,\ldots,N\} \setminus \mathcal{A} \tag{1}$$

as a function of the parameter vector θ. Note that the set $\mathcal{A}$ here is supposed to be a subset of the index set $\{1,2,\ldots,N\}$. This is also referred to as "learning the parameters" or "training the model". Once we are done training a model we can evaluate its classification success rate (CSR) or accuracy as the rate of successfully classified images, by applying it to data we have not used to train, yet know the ground-truth of.

Several notable machine learning textbooks highlight the many issues that can arise during training, e.g. Ref. [29]. It should be clear however, that the model used for prediction can in principle be arbitrarily complex. One way to reduce the complexity is to constrain the possible models to a specific class that can be parametrized by a reasonable number of variables.

2.2 Convolutional Neural Networks

The most popular of such models for image classification are convolutional neural networks (CNNs) [30], which are based on the neuroscientific theory from Ref. [31], in which the authors suggest that local features in an area called the receptive field are detected in early visual layers of the visual cortex and are then progressively combined to create more complex patterns in a hierarchical manner. In the computational implementation of

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

these networks the local features then form the basis upon which we classify the images into defect and non-defect cases.

In many aspects a CNN is a black box from a metrologist's perspective, i.e., the exact way of how it determines which image presents a defect is not easy to understand. However, each of the basic building blocks allows for a intuitive interpretation of how it acts and extracts abstract features on an individual image. The fundamental building blocks of a CNN are i) convolution layers, ii) pooling layers, and iii) fully connected layers, each having a different task in the classification.

The basic layout of a convolution layer is presented in Fig. 2. Each filter or kernel is convoluted with the input volume consisting of d channels and as a result produces a so-called feature map that ideally corresponds to activations that detect specific features of the different classes of images. The user must specify the number of kernels p, their dimension n, m, and the applied stride Δ_s, i.e., the movement of the kernels, when setting up the CNN. The number of floated parameters in this layer then is $(n \cdot m \cdot d) \cdot p + p$, with the additional p variables accounting for the biases on each filter.

The second important layers are pooling layers that provide a form of non-linear down-sampling. They partition the input volume into rectangles, and for each of those rectangles output a value that depends on the values of the entries in those rectangles, usually the maximum. For the pooling layers, the user must specify the size of the rectangles, the applied stride and the method of pooling.

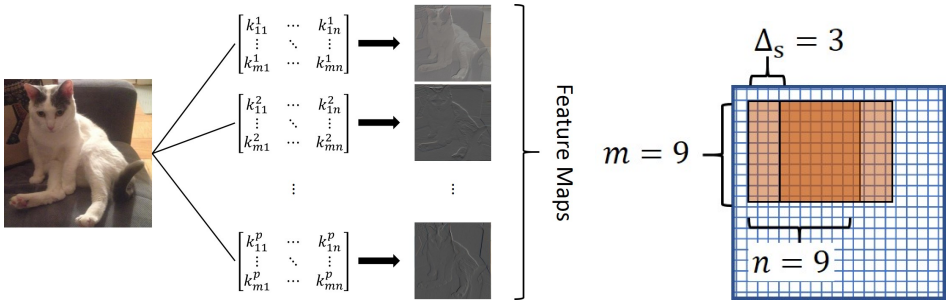


Figure 2. Shematic representation of convolution of an image with a kernel, parameters to be specified are the number of kernels p, their sizes n, m, and the stride Δ_s.

We finally can apply a layer that is fully connected to the previous layer, such that each of its neurons is an affine transformation of the activations of the previous layer, which is obtained by flattening the different activation maps into a single column vector of size n_f. Once we determine the size d of this fully connected layer, we add $d \cdot n_f + d$ variables to the CNN.

Each of the feature maps obtained by applying any of the above mentioned layer types are to be treated with nonlinear activation functions intended to make our model more flexible. In addition there are multiple ways to increase the performance of the classification algorithm, such as dropout layers and batch normalization, see Refs. [32, 33] for details.

2.3 A Base Model

Utilizing an example from the literature allows a closer look at how such models may be improved. The basic model architecture that we want to use for defect detection is taken from [28], which for simplicity we name the "Base" model. Here, the CNN was trained on a database of photometric stereo images of metal surface defects, i.e., rail defects. Each stereo image consisted of 16×16 pixels and $n_c = 2$ channels. The trained CNN outperformed the previous algorithms that used a model-based approach to detect the defects, even without applying regularization, yielding a CSR of around 0.91. The basic layout of the CNN can be seen in Fig. 3 below, the corresponding layer parameters can be found in Tab. 1.

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

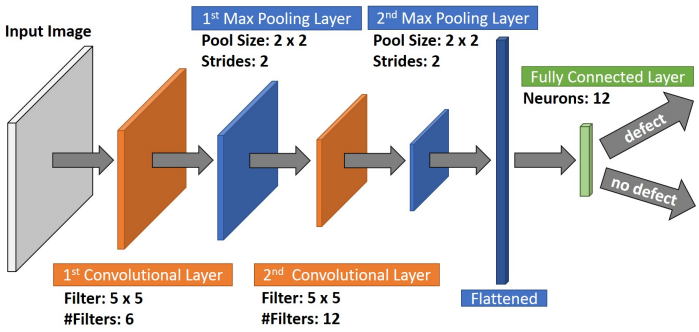


Figure 3. Basic network architecture from [28].

Table 1. Parameters of the Base CNN, comprised of convolutional layers (CL), pooling layers (PL) and the fully connected layer (FCL). The number of total trainable parameters is 6386 (additional parameters are due to applied batch normalization and the final classification).

	1st CL	1st PL	2nd CL	2nd PL	FCL
#Filters	6	N/A	12	N/A	N/A
(Pool,Filter) Size	5× 5	2×2	5×5	2×2	12
Strides	1	2	1	2	N/A
#Parameters	156	0	1812	0	4332

3. OPTIMIZATION FOR DEFECT METROLOGY

3.1 Optimizing the CNN architecture

We based our CNN design on the previous example, with the difference of using two fully connected layers. In order to determine the optimal design, we float several parameters of the CNN in a total of 108 ways, such as the filter sizes at the two convolutional layers $k_1, k_2 \in \{4, 5, 6\}$, and the number of neurons in the first fully connected layer after the flattened layer $d_1 \in \{6, 12, 24, 48\}$. For each value of d_1 we choose three values for the number of neurons in the second fully connected layer $d_2 \in \{0, \frac{d_1}{4}, \frac{d_1}{2}\}$, with $d_2 = 0$ implying that we remove the second fully connected layer, see Fig. 4.

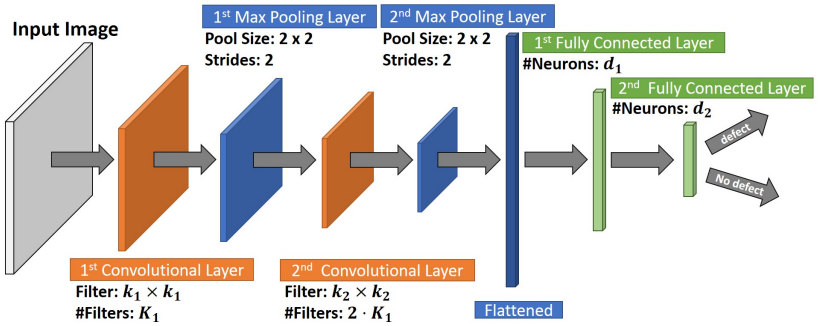


Figure 4. Schematic of modified CNN and the corresponding floated parameters.

We run 30 independent trainings for each of the 108 parameter combinations, each with a fixed learning rate. The resulting CSRs for the different parameter combinations are presented as a boxplot in Fig. 5. One can identify two types of behavior in the plot. Some of the boxplots - that show the 25th percentile, the median, and the 75th percentile - for the CSRs are very tightly centered around the median value, while others spread out more. If one takes a closer look at the parameter combinations that correspond to those tightly centered one finds that they represent models that do not employ a second fully connected layer, i.e., models that have fewer floated parameters if compared to the spread out ones. The reason for this lies in the fact that training models

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

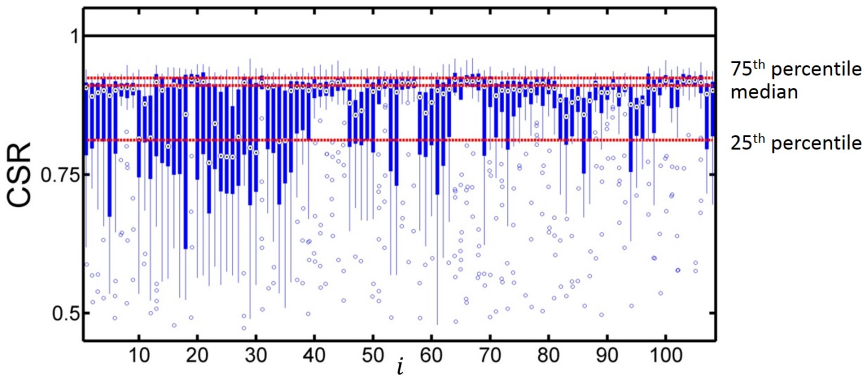


Figure 5. Box plot of classification success rates for different parameter combinations indexed by i. Red dotted lines represent (from top to bottom) the 75th percentile, the median, and 25th percentile of the CSRs for the base model.

that have more parameters can lead to two problems and hence lower CSRs. First, a failed optimization in the sense that due to the immense number of floated parameters we have not found the optimal combination yet, and second overfitting of the model, which means that the model is too flexible and therefore performs perfect on the training data but is not capable of generalizing to previously unseen data [34]. It is also interesting to see that we find a combination of parameters that leads to a model that performs better on the data that the Base model, namely by setting $k_1 = 5, k_2 = 5, d_1 = 48$ and $d_2 = 0$, corresponding to $i = 55$ in Fig. 5. Given that the Base model was motivated by finding defects on images that measured 16×16 pixels it is not surprising that the optimal model in our case with images that consist of 27×24 pixels would need more neurons in the first fully connected layer in order to operate successfully.

3.2 Optimizing Performance Measures

Due to the sigmoid function that we put at the end of our CNN, the classification is performed by assigning a number $\hat{y}_p \in [0, 1]$ to each image. Given our nomenclature, this number reflects the probability that the image shows a non defect structure. The final classification involves a decision threshold τ that is usually set to 0.5, i.e., any image with $\hat{y}_p < 0.5$ will be classified as a defect image. It is obvious that the value of τ has a strong influence on the CSR, and furthermore on other performance measures that might be more useful in semiconductor metrology, such as precision and recall, based on the following definitions, with y denoting the ground-truth label and $\hat{y}$ denoting the predicted label.

True positive (TP): $y = \hat{y} = 0$, True negative (TN): $y = \hat{y} = 1$

False positive (FP): $y = 1, \hat{y} = 0$, False negative (FN): $y = 0, \hat{y} = 1$

Precision: $\pi = \frac{\mathsf{TP}}{(\mathsf{TP}+\mathsf{FP})}$, Recall: $\rho = \frac{\mathsf{TP}}{(\mathsf{TP}+\mathsf{FN})}$, Accuracy: $\mathsf{CSR} = \frac{\mathsf{TP}+\mathsf{TN}}{(\mathsf{TP}+\mathsf{TN}+\mathsf{FP}+\mathsf{FN})}$

Too high a value for τ might lead to too cautious an inspection and hence an increase in recall, and a loss in precision. Too small a value for τ might cause too strict a definition of a defect and hence an increase in precision, with almost no recall. Figure 6 presents the effect of different choices for τ on the accuracy, precision, and recall for three different models and also presents the resulting histograms of the $\hat{y}_p$ of the validation set. Even though the CSR for all three models is around 0.91, they behave quite differently. The first model is very uncertain about the class of the images and outputs several values of $\hat{y}_p$ that are close to 0.5, hence the precision, recall and CSR are very sensitive towards changes in the decision threshold τ. The second model yields a better approximation of the bimodal distribution of the labels, however it is not perfect and still shows a significant sensitivity towards changes of τ. The third model's performance is very good as it manages to more accurately reproduce the strictly bimodal distribution of the class labels.

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

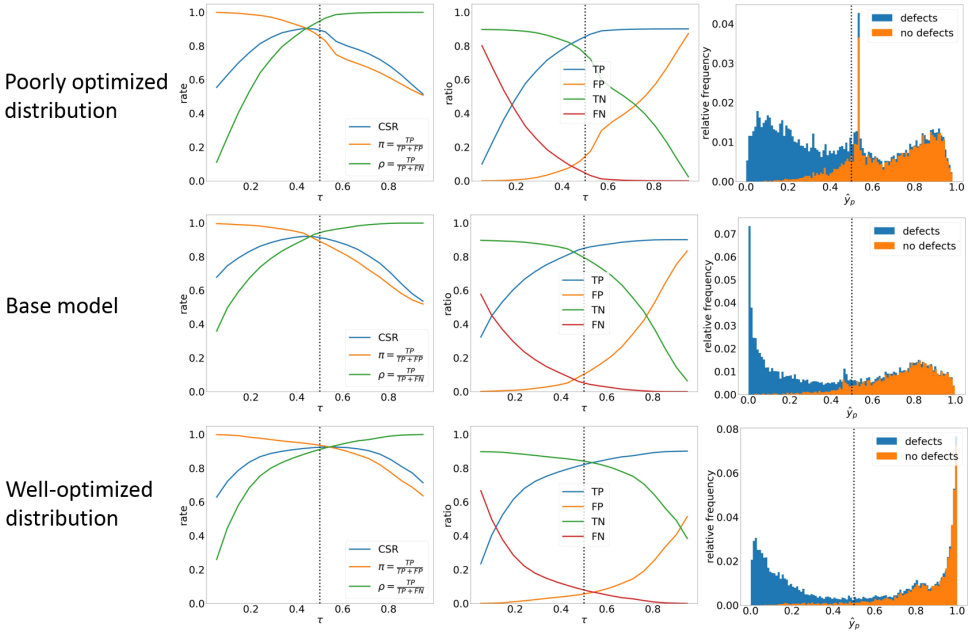


Figure 6. Effect of τ on CSR, π and ρ for three different CNNs. top: $k_1 = k_2 = 4, d_1 = 12, d_2 = 0, i = 1$, middle: the Base model $k_1 = k_2 = 5, d_1 = 12, d_2 = 0, i = 49$, bottom: $k_1 = k_2 = 6, d_1 = 48, d_2 = 24, i = 105$. The CSR does not reveal the nature of the $\hat{y}_p$ distributions.

4. RESULTS AND CONCLUSION

In this work we have presented an approach to defect detection based on convolutional neural networks. We based our research on a previously used CNN that had been designed to detect defects in steel rails. Even though the tasks are quite different, we could achieve a CSR around 0.91 with this model. The intrinsically stochastic nature of the training of the network however makes it hard to evaluate the CNNs' performance based on one realization alone, we therefore varied several of the model parameters and trained the corresponding models multiple times. It turned out that some parameter combinations led to a more stable performance, i.e., the obtained CSRs varied less if compared to the Base model. However, the combinations tested failed to yield a CSR that improved upon the median CSR of the Base model.

CSR is a straightforward metric to evaluate the performance of a classification algorithm, but its inherent design should not discriminate between the relative impact of false positives and false negatives. This general metric may be inappropriate from a metrologists' perspective, since there can be an intrinsic economic cost to false positives that is different from the costs for false negatives that is usually not accounted for in traditional machine learning. We have presented a simple method for adjusting the model to allow one to tune their own threshold of desired FPs and FNs. This simplistic approach can be improved upon either by including cost within the layers or including cost within the loss function itself [35, 36].

It should be noted that the optimization steps were performed using training data from a single defect, a very simplified approach chosen to illustrate our methodology. In practical applications defects can have several shapes and optical scattering signatures. The findings presented above should however be easily extendable to a wider range of defects. Our current research is addressing the question how our model could be extended or adapted to unseen defect types, a problem also referred to as transfer learning [37].

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

References

[1] K. Wells, G. Chen, M. Derstine, S. Lange, and D. Shortt, "Extending optical inspection to the VUV," in 2017 International Conference on Frontiers of Characterization and Metrology for Nanoelectronics (FCMN), D. G. Seiler, ed., pp. 92–101, NIST, (Monterey, CA USA), 2017.

[2] B. M. Barnes, H. Zhou, M.-A. Henn, M. Y. Sohn, and R. M. Silver, "Assessing the wavelength extensibility of optical patterned defect inspection," Proc. SPIE **10145**, p. 1014516, 2017.

[3] B. M. Barnes, H. Zhou, M.-A. Henn, M. Y. Sohn, and R. M. Silver, "Optimizing image-based patterned defect inspection through FDTD simulations at multiple ultraviolet wavelengths," Proc. SPIE **10330**, p. 103300W, 2017.

[4] D. Meshulach, I. Dolev, Y. Yamazaki, K. Tsuchiya, M. Kaneko, K. Yoshino, and T. Fujii, "Advanced lithography: wafer defect scattering analysis at DUV," Proc SPIE **7638**, p. 76380K, 2010.

[5] N. Rana, Y. Zhang, T. Kagalwala, and T. Bailey, "Leveraging advanced data analytics, machine learning, and metrology models to enable critical dimension metrology solutions for advanced integrated circuit nodes," Journal of Micro/Nanolithography, MEMS, and MOEMS **13**(4), p. 041415, 2014.

[6] J. Kim, S. Kim, N. Kwon, H. Kang, Y. Kim, and C. Lee, "Deep learning based automatic defect classification in through-silicon via process: Fa: Factory automation," in 2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), pp. 35–39, IEEE, 2018.

[7] D. S. Woldeamanual, A. Erdmann, and A. Maier, "Application of deep learning algorithms for lithographic mask characterization," Proc. SPIE **10694**, p. 1069408, International Society for Optics and Photonics, 2018.

[8] M.-A. Henn, B. M. Barnes, H. Zhou, and R. M. Silver, "Optimizing defect detectability across multiple ultraviolet wavelengths," Mathematics in Imaging , 2018.

[9] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," IEEE Transactions on Semiconductor Manufacturing **31**(2), pp. 309–314, 2018.

[10] M. Niskanen, O. Silvén, and H. Kauppinen, "Color and texture based wood inspection with non-supervised clustering," Proceedings of the scandinavian Conference on image analysis , 2001.

[11] I.-S. Tsai, C.-H. Lin, and J.-J. Lin, "Applying an artificial neural network to pattern recognition in fabric defects," Textile Research Journal **65**, 1995.

[12] N. Rana and C. Chien, "Deep machine learning based image classification in hard disk drive manufacturing (conference presentation)," Proc. SPIE **10585**, p. 105850Y, 2018.

[13] N. Rana and C. Chien, "Predicting back-end writer main pole shape geometry from wafer metrology data in hard disk drive manufacturing (conference presentation)," Proc. SPIE **10585**, p. 1058515, 2018.

[14] M. Zhang, J. Wu, H. Lin, P. Yuan, and Y. Song, "The application of one-class classifier based on cnn in image defect detection," Procedia computer science **114**, pp. 341–348, 2017.

[15] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," The International Journal of Advanced Manufacturing Technology **94**(9-12), pp. 3465–3471, 2018.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems , 2012.

[17] S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, et al., "A 14nm logic technology featuring 2 nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm 2 SRAM cell size," Electron Devices Meeting (IEDM) , 2014.

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

[18] B. M. Barnes, "Geometries and material properties for simulating semiconductor patterned bridge defects using the finite-difference-time-domain (FDTD) method," http://doi.org/10.18434/T4/1500937 .

[19] B. M. Barnes, M. Y. Sohn, F. Goasmat, H. Zhou, A. E. Vladár, R. M. Silver, and A. Arceo, "Three-dimensional deep sub-wavelength defect detection using λ= 193 nm optical microscopy," Optics express , 2013.

[20] B. M. Barnes, F. Goasmat, M. Y. Sohn, H. Zhou, A. E. Vladár, and R. M. Silver, "Effects of wafer noise on the detection of 20-nm defects using optical volumetric inspection," Journal of Micro/Nanolithography, MEMS, and MOEMS , 2015.

[21] A. Taflove, "Application of the finite-difference time-domain method to sinusoidal steady-state electromagnetic-penetration problems," IEEE Transactions on electromagnetic compatibility , 1980.

[22] R. Durrett, Probability: Theory and Examples, Cambridge university press, 2010.

[23] R. M. Silver, B. M. Barnes, Y. Sohn, R. Quintanilha, H. Zhou, C. Deeb, M. Johnson, M. Goodwin, and D. Patel, "The limits and extensibility of optical patterned defect inspection," Proc. SPIE **7638**, p. 76380J, 2010.

[24] B. M. Barnes, R. Quinthanilha, Y.-J. Sohn, H. Zhou, and R. M. Silver, "Optical illumination optimization for patterned defect inspection," Proc. SPIE **7971**, p. 79710D, 2011.

[25] B. M. Barnes, F. Goasmat, M. Y. Sohn, H. Zhou, R. M. Silver, and A. Arceo, "Enhancing 9 nm node dense patterned defect optical inspection using polarization, angle, and focus," Proc. SPIE **8681**, p. 86810E, 2013.

[26] B. M. Barnes, M. Y. Sohn, F. Goasmat, H. Zhou, A. E. Vladár, R. M. Silver, and A. Arceo, "Three-dimensional deep sub-wavelength defect detection using λ= 193 nm optical microscopy," Optics express **21**(22), pp. 26219–26226, 2013.

[27] B. M. Barnes, F. Goasmat, M. Y. Sohn, H. Zhou, A. E. Vladár, and R. M. Silver, "Effects of wafer noise on the detection of 20-nm defects using optical volumetric inspection," Journal of Micro/Nanolithography, MEMS, and MOEMS **14**(1), p. 014001, 2015.

[28] D. Soukup and R. Huber-Mörk, "Convolutional neural networks for steel surface defect detection from photometric stereo images," International Symposium on Visual Computing , 2014.

[29] E. Alpaydin, Introduction to machine learning, MIT press, 2009.

[30] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, MIT press Cambridge, 2016.

[31] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," The Journal of physiology , 1962.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research , 2014.

[33] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," International Conference on Machine Learning , 2015.

[34] D. M. Hawkins, "The problem of overfitting," Journal of chemical information and computer sciences , 2004.

[35] Y. Zhang and Z.-H. Zhou, "Cost-sensitive face recognition," IEEE transactions on pattern analysis and machine intelligence **32**(10), pp. 1758–1769, 2010.

[36] Y.-A. Chung, G.-H. Lee, and S.-W. Yang, "Cost-sensitive classification with deep learning using cost-aware pre-training," Mar. 9 2017. US Patent App. 14/757,959.

[37] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on knowledge and data engineering **22**(10), pp. 1345–1359, 2010.

Henn, Mark Alexander; Zhou, Hui; Silver, Richard; Barnes, Bryan.
"Applications of machine learning at the limits of form-dependent scattering for defect metrology."
Paper presented at ADVANCED LITHOGRAPHY 2019: TECHNOLOGIES FOR LITHOGRAPHY R&D, DEVICES, TOOLS, FABRICATION, AND SERVICES, San Jose, CA, United States. February 24, 2019 - February 28, 2019.

Discriminated properties of PEI functionalized gold nanoparticles (Au-PEIs) predetermined by synthetic routes of ligand exchange and reduction processes: Paths and Fates

Tae Joon Cho*, John M. Pettibone, and Vincent A. Hackley

Material Measurement Laboratory, National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, MD 20899
*email: taejoon.cho@nist.gov

ABSTRACT

Polyethyleneimine (PEI) functionalized AuNPs (Au-PEIs) have potential use as positively charged gold nanoparticles (AuNPs) for nano-medicinal applications, due to their cationic surfaces that promote cellular uptake and gene transfection.[1-6] Au-PEIs can be prepared by one of two methods: (1) ligand modification (exchange or co-attachment) of AuNPs and PEI, or (2) reduction of Au^{III} ions in the presence of PEI. Herein, the ligand exchanged Au-PEIs (Au-PEI[*LE*]s) and the materials by reduction (Au-PEI[*Red*]s) were prepared from citrate-stabilized AuNPs (10 nm, nominal) and $HAuCl_4$, respectively. We demonstrate differentiated product formation via each synthetic route and their discriminated physico-chemical properties by systematic examination. The physico-chemical properties and the conjugation mechanisms of Au-PEI[*LE*] and Au-PEI[*Red*] were characterized by orthogonal analyses system. Furthermore, the colloidal stabilities of all produced Au-PEIs were investigated under various physiologically relevant conditions and shown that the consequential fates – property and colloidal stability – of Au-PEIs were dependent on born pathways significantly. We found that some of Au-PEIs showed outstanding colloidal stability in certain circumstances which is very critical to be a drug carrier in nanomedicine.

Keywords: gold nanoparticles, polyethyleneimine, PEI, gold PEI conjugates, positively charged, synthesis, characterization, stability

1 INTRODUCTION

Polyethyleneimine (PEI) coated gold nanoparticles (AuNPs), hereafter Au-PEIs, have been studied due to several biologically relevant features, including interaction with anionic polyelectrolytes[7] and biological entities (cell surface),[8] cellular uptake,[9-11], gene delivery/transfection,[1-6, 12] hydrophilicity, and proton scavenging (pH tolerance).[6, 8] In general, Au-PEIs can be prepared by *ligand exchange* of surface modified AuNPs by PEI chains (Au-PEI[*LE*]), and *reduction method* started from $HAuCl_4$ in the presence of PEI (Au-PEI[*Red*]). Despite the breadth of the aforementioned work, it is still unclear how the specific properties of PEI and the methods utilized to conjugate it to AuNPs impact the final product size and stability (which presumably impact performance). Furthermore, knowledge regarding the stability of Au-PEIs under physiologically relevant conditions is critically important for biological applications, yet this has not generally been reported in prior studies. Colloidal stability can dramatically impact the findings for bio-relevant research, where performance can be impacted by one distinct population within a polydisperse material, agglomeration/aggregation effects, or surface modifications.[13] Here in, we explored a range of synthetic routes by systematic examination of variables, including reaction condition, the molar mass/backbone structure of PEIs, and molar mass ratio (PEI:Au). Au-PEI products were evaluated by a variety of approaches to evaluate the impact of synthetic route on physico-chemical properties. Orthogonal analytical methods were utilized including dynamic light scattering (DLS), ultraviolet-visible (UV-Vis) spectroscopy, transmission electron microscope (TEM), and attenuated total reflectance FT-IR (ATR-FTIR) measurements. Additionally, the colloidal stability for each produced Au-PEIs was investigated under physiologically relevant conditions including shelf-life (at least 6 months period), temperature variation, wide range of pH values, and behaviors in biological level media. We demonstrated that some of Au-PEIs showed remarkable colloidal stability which is a key requirement for drug carrier candidates in nanomedicine. In closing, comparing physico-chemical properties of Au-PEI produced between different synthetic approaches can provide the nanomedicine community useful information for better design of nano-vector development.

2 EXPERIMENTS

2.1 Materials and Instruments[ξ]

$Gold^{III}$ chloride hydrate ($HAuCl_4 \cdot 3H_2O$, ACS reagent) was purchased from Sigma-Aldrich (St. Louis, MO). Branched 2 kDa PEI (PEI2kB, 50 % mass fraction in water)

[ξ] The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology.

Cho, Tae Joon; Pettibone, John; Hackley, Vincent.
"Discriminated properties of PEI functionalized gold nanoparticles (Au-PEIs) predetermined by synthetic routes of ligand exchange and reduction processes: Paths and Fates."
Paper presented at TechConnect World Innovation Conference 2019, Boston, MA, United States. June 17, 2019 - June 19, 2019.

and 25 kDa PEI (PEI25kB) were obtained from Sigma-Aldrich (St. Louis, MO). 10 kDa branched PEI (PEI10kB) and 25 kDa linear PEI (PEI25kL) were obtained from Polysciences Inc. (Warrington, PA). Other specific reagents used in this study are identified in the references.[14, 15] All chemicals were used without further purification. Deionized water (18.2 MΩ·cm) was produced by an Aqua Solutions (Jasper, GA) Type I biological grade water purification system. Details regarding other instruments (DLS, UV-Vis, TEM, and ATR-FTIR) and methodology are also provided in references.[14, 15] The uncertainty of size and zeta potential represent the mean and one standard deviation of at least three measurements.

2.2 Preparation of Au-PEIs

General method for **Au-PEI*[LE]*s**

One mL of each PEI solution (1.0 w.% in DI water) was added dropwise to 10 mL of citrate-stabilized gold colloid suspension, which was then stirred at room temperature for 5 h and purified using centrifugal filtration (Amicon Ultra, Millipore; regenerated cellulose membrane, MWCO = 100 kDa). The filtrate was removed and re-diluted with DI water to the starting concentration, and then purified Au-PEIs were passed through a 0.2 µm nylon filter to remove any large impurities such as dust. It is important to note that the purification step was not applied to Au-PEI2kB and Au-PEI25kL due to their rapid aggregation during the conjugation process.

General method for **Au-PEI[*Red*]s**

One ml of each PEI solution (PEI25kL; 1.0 w.%, branched PEIs; 10 w.% in DI water) was added to a 10 mL of 2.5 mmol/L of aqueous $HAuCl_4$ solution in a borosilicate glass vial covered with a cap (EPA vials, Thermo Scientific, Waltham, MA) at room temperature and heated up to 80 °C using a magnetic hot plate with stirring. In these experiments, the temperature was increased from room temperature at a rate of about 5 °C/min. After reaching 80 °C, the reaction mixture was stirred for 1.5 h, then cooled down (removal from hotplate) to the ambient temperature. Resulting products were purified by dialysis (molecular weight cut off 100 kDa) against DI water for 2 days.

3 RESULTS AND DISCUSSIONS

3.1 Synthesis and Characterization

When preparing Au-PEI[*LE*]s, PEI stocks are introduced to the citrate AuNP suspension, yielding a slight color change (ruby red to purplish red) after reaction, indicative of the ligand exchange from citrate to PEI. However, in case of linear PEI (PEI25kL), the reaction mixture was precipitated within 5 min during the reaction (aggregation). The final product by reduction method yielded a ruby red suspension of Au-PEI[*Red*]s when using branched PEIs. In contrast, employing linear PEI yielded a cloudy pale red violet color suspension. The physico-chemical properties of Au-PEIs were determined by orthogonal measurement techniques including DLS (z-average hydrodynamic diameter (D_H), zeta potential (ZP)), TEM (D_{TEM}), and UV-Vis (surface plasmon resonance (SPR) band). Representative data of each Au-PEIs were presented in Table 1.

Table 1. Fundamental properties of Au-PEIs

Au-PEIs	D_H[a] (nm)	D_{TEM}[b] (nm)	SPR[c] (nm)	ZP[d] (mV)
Citrate AuNPs	12.1 ± 0.1	8.3 ± 1.3	518	-37.1 ± 1.4
Au-PEI2kB[*LE*]	21.8 ± 0.7	8.3 ± 0.8	524	+24.6 ± 3.0
Au-PEI10kB[*LE*]	62.9 ± 0.3	8.4 ± 0.7	530	+25.8 ± 0.7
Au-PEI25kB[*LE*]	78.8 ± 0.6	8.4 ± 0.9	534	+29.4 ± 1.3
Au-PEI25kL[*LE*]	N/A[e]	N/A[e]	N/A[e]	N/A[e]
Au-PEI2kB[*Red*]	21.5 ± 0.5	11.3 ± 2.0	522	+12.7 ± 1.4
Au-PEI10kB[*Red*]	22.0 ± 0.3	8.1 ± 2.6	523	+21.2 ± 1.2
Au-PEI25kB[*Red*]	22.2 ± 0.4	8.6 ± 2.3	524	+22.5 ± 1.7
Au-PEI25kL[*Red*]	88.2 ± 0.7	30.3 ± 7.0	545	+40.0 ± 0.4

a: D_H obtained by DLS, *b*: Au core size obtained by TEM; more than 150 particles were measured for each Au-PEIs and size reported is the population mean with one standard deviation about the mean., *c*: surface plasmon resonance band by UV-Vis measurements, *d*: zeta potential by DLS, and *e*: not available due to the precipitation during the reaction.

As shown in Table 1, D_H and SPR bands (by UV-Vis) of the Au-PEI[*LE*]s showed significant changes when compared to the starting citrate AuNP, increasing in value with the molar mass of associated PEIs (branched), while the core sizes (D_{TEM}) were almost identical regardless of PEIs. On the other hand, synthesizing Au-PEI[*Red*]s using branched PEI yield nearly identical values of D_H, D_{TEM}, and SPR for each Au-PEIs. These results clearly demonstrate that the application of different synthetic pathways result in Au-PEIs with varying properties. ZPs of Au-PEIs indicate the successful creation of positively charged AuNPs in comparison of negative surface of citrate AuNPs.

3.2 Reaction mechanism of Au-PEIs

ATR-FTIR studies have been conducted to determine if the synthetic mechanism of Au-PEI[*LE*]s follows a direct ligand exchange or layer *by* layer (L-*b*-L) process. To examine the interaction between PEI and citrate moiety on AuNPs, we conducted flow cell experiments with ATR-FTIR (Figure 1).[15] The ATR-FTIR spectra for the addition of PEI2kB to the citrate-stabilized AuNP film is shown in spectrum (i). The spectrum after the addition of three 1 mL aliquots of 1.0 % mass fraction solution of PEI over a similar mixing time for the general method is shown in spectrum (iii).

Absorbance has been normalized and spectra offset on

the vertical axis for ease of viewing. The data clearly shows the loss of each characteristic citrate vibrational mode after the addition of the PEI2kB. Although the amount of citrate remaining was not quantitatively determined, the resulting spectrum for the Au-PEI2kB film after flushing and drying

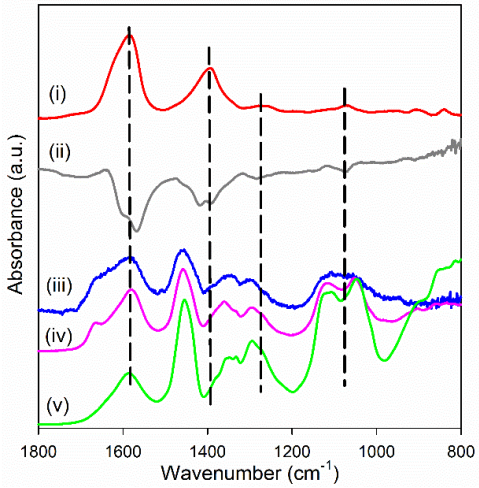
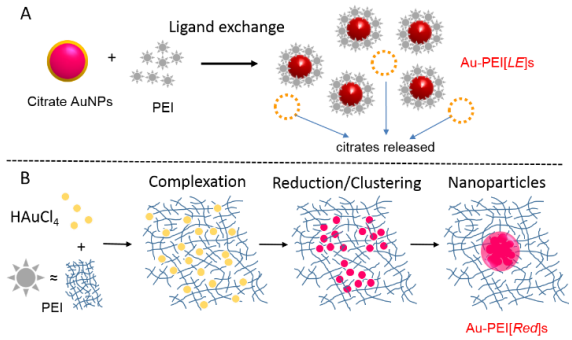


Figure 1. In-situ (in the flow cell) ATR-FTIR spectra for (i) citrate-stabilized AuNPs, (ii) loss spectrum after the addition of PEI2kB in the flow cell, (iii) dried and washed product after PEI2kB addition, and reference spectrum of (iv) dried Au-PEI2kB conjugate (separate batch, purified) and (v) dried PEI2kB (only) for comparison.

exhibits predominantly bands associated with PEI. The removal of citrate is further supported by the comparison of the dried film after rinsing (spectrum iii) and the purified separate batch sample (spectrum iv) that exhibit very similar signatures with predominantly PEI vibrations. These serial results obtained in situ support the replacement of citrate moieties by PEI chains. Consequently, the ATR spectra provide evidence that conjugation is principally accomplished by a ligand exchange reaction rather than an electrostatic-induced L-*b*-L process between negative surface (citrate) and positively charged ligand (PEI).



Scheme 1. Illustrative presentation of synthetic process of Au-PEIs depends on preparation route; a) by ligand exchange, and b) reduction method.

Based on the results as described above, we proposed the synthetic mechanisms of Au-PEIs in Scheme 1. Panel A illustrates the ligand exchange process that citrate moieties are replaced with PEI chains on gold core surface and panel B demonstrates reduction process of Au^{III} ions with PEIs and formation of the final Au^0 nanoparticles.

3.3 Stability study

Colloidal stability is an important issue for any application of AuNPs. We evaluated the stability of the Au-PEIs over a range of relevant conditions utilizing previously established protocols[16] and compared those of as functions of molar mass of PEI and synthetic pathways in both. For the study of long-term stability, Au-PEIs aged for 6 months under ambient laboratory conditions yielded a size distribution and SPR band (Figure 2).

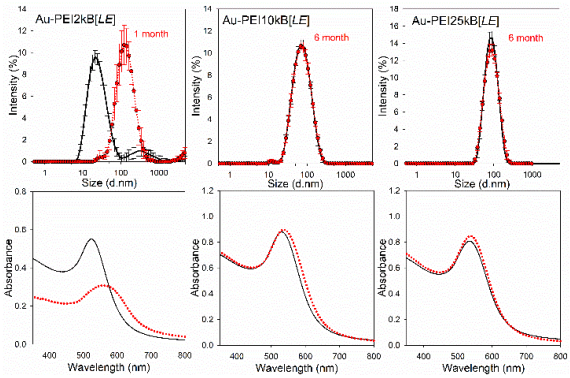


Figure 2. Shelf-life test of Au-PEI[*LE*]s; (upper) DLS size distributions for the initial product (black line) and after aging (red line), (bottom) UV-Vis absorbance showing SPR band for the initial product (black line) and after aging (red line). Error bars represent standard deviation.

The aged samples of Au-PEI10kB and -25kB yielded nearly identical DLS mean size and size distributions compared with the initial, freshly prepared and purified products (Figure 2, top). In contrast, those for Au-PEI2kB significantly increased within one month, indicating reduced colloidal stability compared with the higher molar mass PEIs. The associated SPR band intensities and peak positions (Figure 2, bottom) provide a more sensitive probe to changes in the near field environment of the AuNPs.[15] Notably, the degree of long-term stability and PEI molar mass dependency were very similarly observed in Au-PEI[*Red*]s (data omitted).

For biological application, stability in physiological media is critical. Based on a previous study,[14] citrate-stabilized AuNPs showed immediate instability in phosphate buffered saline (PBS), otherwise Au-PEIs exhibited improved stability in PBS over 48 h period relevant to cell exposure assays (Figure 3). In the long-term stability study, molar mass of PEI impacts the colloidal stability of Au-PEIs[15] as shown in figure 2. Figure 3

Cho, Tae Joon; Pettibone, John; Hackley, Vincent.
"Discriminated properties of PEI functionalized gold nanoparticles (Au-PEIs) predetermined by synthetic routes of ligand exchange and reduction processes: Paths and Fates."
Paper presented at TechConnect World Innovation Conference 2019, Boston, MA, United States. June 17, 2019 - June 19, 2019.

demonstrates how significant it is against physiological environment. Furthermore, the effect of synthetic pathways to the physico-chemical properties of resulted Au-PEIs was clearly observed by this study. The overall stability of Au-PEI[Red]s exhibited significant improvements compared to those of Au-PEI[*LE*]s. Based on these results, we select Au-PEI25kB[*Red*]s, as the most stable Au-PEI, and conducted

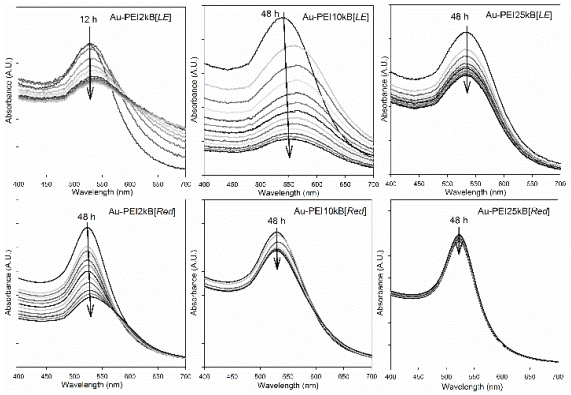


Figure 3. Stability of Au-PEIs in PBS over time, as monitored by UV-Vis: (Top) Au-PEI[*LE*]s, and (Bottom) Au-PEI[*Red*]s; PEI2kB, -10kB, and -25kB, from left to right in both panels.

further colloidal stability tests. Over a wide range of pH values (pH 1.5 ~ 12), Au-PEI25kB[*Red*]s showed remarkable stability over 12 h (data omitted). Overall, the resistance against acid destabilization is greatly improved relative to citrate AuNPs and all of the remaining Au-PEI synthetic conditions. Thermal stability of Au-PEI25kB[*Red*]s was evaluated by UV-Vis from (20 to 60) °C, a relevant temperature range for most biological assays. The constancy of the SPR band (from UV-Vis spectra, data omitted) confirm that the Au-PEI25kB[*Red*]s are stable with respect to temperature variations over the tested range.

4 CONCLUSION

The positively charged gold nanoparticles, Au-PEIs were prepared via two different pathways; 1) ligand modification, and 2) reduction process. In summary, we demonstrated that each synthetic process resulted in differentiated product formation and their physico-chemical properties, by systematic examination considering reaction condition and the dependency of relative molar mass/backbone structure of PEIs. The physico-chemical properties and the conjugation mechanisms of Au-PEI[*LE*] and Au-PEI[*Red*] were characterized by an orthogonal analyses system including DLS, UV-Vis, TEM, and ATR-FTIR measurements. Furthermore, the colloidal stabilities of all produced Au-PEIs were investigated under various physiologically relevant conditions. We found that the consequential fates of Au-PEIs were significantly dependent on born pathways. Notably, Au-PEI25kB[*Red*]s showed the most remarkable stability in such circumstances, which is very critical to be applied in nanomedicine and nanobiotechnology. We are currently working on this material for advanced studies including hybridization with biological entities to be gene transfection tools.

REFERENCES

[1] Y. Ding, Z. W. Jiang, K. Saha, C. S. Kim, S. T. Kim, R. F. Landis and V. M. Rotello, *Mol. Therapy*, 22, 1075, 2014.

[2] A. Elbakry, E. C. Wurster, A. Zaky, R. Liebl, E. Schindler, P. Bauer-Kreisel, T. Blunk, R. Rachel, A. Goepferich and M. Breunig, *Small*, 8, 3847, 2012.

[3] Z. Z. Chen, L. F. Zhang, Y. L. He and Y. F. Li, *Acs Appl. Mat. Interfaces*, 6, 14196, 2014.

[4] M. Y. Lee, S. J. Park, K. Park, K. S. Kim, H. Lee and S. K. Hahn, *Acs Nano*, 5, 6138, 2011.

[5] W. J. Song, J. Z. Du, T. M. Sun, P. Z. Zhang and J. Wang, *Small*, 6, 239, 2010.

[6] M. Thomas and A. M. Klibanov, *Proc. Nat. Acad. Sci., U.S.A.*, 100, 9138, 2003.

[7] G. Kramer, H. M. Buchhammer and K. Lunkwitz, *Coll. Surf. A-Physicochem. Engineer. Aspects*, 137, 45, 1998.

[8] O. Boussif, F. Lezoualch, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix and J. P. Behr, *Proc. Nat. Acad. Sci., U.S.A.*, 92, 7297, 1995.

[9] E. C. Cho, J. W. Xie, P. A. Wurm and Y. N. Xia, *Nano Letters*, 9, 1080, 2009.

[10] U. Taylor, S. Klein, S. Petersen, W. Kues, S. Barcikowski and D. Rath, *Cytometry Part A*, 77A, 439, 2010.

[11] R. R. Arvizo, O. R. Miranda, M. A. Thompson, C. M. Pabelick, R. Bhattacharya, J. D. Robertson, V. M. Rotello, Y. S. Prakash and P. Mukherjee, *Nano Letters*, 10, 2543, 2010.

[12] P. Zhang, B. B. Li, J. W. Du and Y. X. Wang, *Coll. Surf. B-Biointerfaces*, 157, 18, 2017.

[13] R. I. MacCuspie, *J. Nanopart. Res.*, 13, 2893, 2011.

[14] T. J. Cho, R. I. MacCuspie, J. Gigault, J. M. Gorham, J. T. Elliott and V. A. Hackley, *Langmuir*, 30, 3883, 2014.

[15] T. J. Cho, J. M. Pettibone, J. M. Gorham, T. M. Nguyen, R. I. MacCuspie, J. Gigault and V. A. Hackley, *Langmuir*, 31, 7673, 2015.

[16] T. J. Cho and V. A. Hackley, National Institue of Standards and Technology https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1200-26.pdf, 2018.

Cho, Tae Joon; Pettibone, John; Hackley, Vincent.
"Discriminated properties of PEI functionalized gold nanoparticles (Au-PEIs) predetermined by synthetic routes of ligand exchange and reduction processes: Paths and Fates."
Paper presented at TechConnect World Innovation Conference 2019, Boston, MA, United States. June 17, 2019 - June 19, 2019.

Manipulation Data Collection and Annotation Tool for Media Forensics

Eric Robertson[1], Haiying Guan[2], Mark Kozak[1], Yooyoung Lee[2], Amy N. Yates[2], Andrew Delgado[2], Daniel Zhou[2], Timothee Kheyrkhah[2], Jeff Smith[3], Jonathan Fiscus[2]

[1]PAR Government, [2]National Institute of Standards and Technology, [3]University of Colorado Denver

eric_robertson@partech.com, haiying.guan@nist.gov, mark_kozak@partech.com, yooyoung.lee@nist.gov, amy.yates@nist.gov, andrew.delgado@nist.gov, daniel.zhou@nist.gov, timothee.kheyrkhah@nist.gov, jeff.smith@ucdenver.edu, jonathan.fiscus@nist.gov

Abstract

With the increasing diversity and complexity of media forensics techniques, the evaluation of state-of-the-art detectors are impeded by lacking the metadata and manipulation history ground-truth. This paper presents a novel image/video manipulation Journaling Tool (JT) that automatically or semi-automatically helps a media manipulator record, or journal, the steps, methods, and tools used to manipulate media into a modified form. JT is a unified framework using a directed acyclic graph representation to support: recording the manipulation history (journal); automating the collection of operation-specific localization masks identifying the set of manipulated pixels; integrating annotations and metadata collection; and execution of automated manipulation tools to extend existing journals or automatically build new journals.

Using JT to support the 2017 and 2018 Media Forensics Challenge (MFC) evaluations, a large collection of image manipulations was assembled that included a variety of different manipulation operations across image, video, and audio. To date, the MFC's media manipulation team has collected more than 4500 human-manipulated image journals containing over 100,000 images, more than 400 manipulated video journals containing over 4,000 videos, and generated thousands of extended journals and hundreds of auto-manipulated journals.

This paper discusses the JT's design philosophy and requirements, localization mask production, automated journal construction tools, and evaluation data derivation from journals for performance evaluation of media forensics applications. JT enriches the metadata collection, provides consistent and detailed annotations, and builds scalable automation tools to produce manipulated media, which enables the research community to better understand the problem domain and the algorithm models.

1. Introduction

Media forensics is the science and practice of determining the authenticity and establishing the integrity of an audio and visual media asset [1][2][3][4][5][6][7][8] for a variety of use cases such as litigation, fraud investigation, etc. For computer researchers, media forensics is an interdisciplinary approach to detect and identify digital media alterations using forensic techniques based on computer vision, machine learning, media imaging, statistics, etc. to identify evidence (or indicators) supporting or refuting the authenticity of a media asset.

Existing media manipulation detection technologies forensically analyze media content for indicators using a variety of information sources and techniques such as the Exchangeable Image File Format (EXIF) header, camera Photo Response Non Uniformity (PRNU) model, manipulation operation (e.g. splice, copy-clone) detection, compression anomalies, and physics-based and semantic-based consistency approaches. Before the JT, there was no manipulation annotation tool capable of capturing a historical record of the manipulation and related metadata to enable the evaluation of a wide variety of technologies and specific aspects of technologies from a single journal.

1.1. Background

DARPA's Media Forensics (MediFor) program [9] brings together world-class researchers to develop technologies for the automated integrity assessment of a media in an end-to-end platform. The primary objective for the MediFor data collection and evaluation team is to create benchmark datasets that advance current technologies and drive technological developments by understanding the key factors of this domain. The data collection and media manipulation team provides various kinds of data, metadata, and annotations supporting the program evaluations, while the evaluation team designs the data collection requirements, validates the quality of the collected data, and assembles the evaluation datasets.

1.2. Related work

The media forensics and anti-forensics techniques are developing quickly in recent years, but the evaluation and analysis of state-of-the-art manipulation detectors are impeded by the diversity and complexity of the technologies and the limitations of existing datasets, which include but are not limited to: (i) lack of rich metadata (annotations) essential to systematic evaluations and analysis; (ii) missing structured representation of manipulation history reconstruction; (iii) insufficient detail to generate diverse evaluation metadata and ground-truth (e.g. image format, manipulation semantic meaning, camera information, and manipulated image masks) for specific detectors given the same manipulated media (image or video). For a summary of existing media forensics datasets, please refer to Section 2 in [10] for details.

The ultimate goal of the MediFor program is to gain a

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

deep understanding of the performance of different technologies based on the properties of the media, their manipulations, and their relationships with each other. In order to meet this goal, the program requires a large amount of highly diverse imagery with ground-truth labels and metadata covering an enormous spectrum of media itself, and manipulation types from the diverse image editing software and tools.

A thorough understanding of algorithm performance requires analysis of multiple factors that go into the production of manipulated imagery. These factors include the steps to produce the manipulation, software used for the manipulation, parameters provided to the software during the manipulation, and anti-forensics to disguise the manipulation. A complete assessment of state-of-the-art detectors using factor analysis with detail metadata annotations provides vital information for further advancement of current technologies. However, data collection, manipulation, and annotation are all labor intensive. In our initial manipulation and annotation study, the time used to perform the manipulation (splice, clone, or remove) is nearly the same as the time used to annotate the metadata. Our objective is to develop a tool to assist manipulators annotate data efficiently and effectively.

Computer vision researchers have developed many tools to help them collect the research data and label ground-truth, such as Photostuff [11], LabelMe [12], VATIC (Video Annotation Tool from Irvine California) [13] for the VIRAT dataset, Computer Vision Annotation Tool (CVAT) [14], VGG Image Annotator [15], Scalabel [16] and BeaverDam [17] for UC Berkeley's DeepDrive project, Polygon-RNN++ [18], and Microsoft Visual Object Tagging Tool (VoTT) [19]. Google recently announced their new 'Google Cloud Video Intelligence API' [20], which uses machine learning and the cloud to automatically analyze and annotate video content. VideoTagger [21] is another annotation tool for biological study.

Different tools are designed for different applications in different research domains. The existing annotation tools are not suitable for media forensic annotation for several reasons. (1) Most existing tools label data using annotators' basic knowledge, such as labeling an object class or segment an object out given an image. But in media forensics, given an image, the annotator may not know for sure if it was manipulated, or which pixels were changed. Such information cannot be easily recovered after the manipulation was done (see Section 4.2 for explanation and example). Therefore, annotation tools that document after the media has already been processed are not suitable for media forensic annotation. (2) Existing annotation tools do not document a thorough trace of manipulation steps. In-depth traces are needed because: Different manipulation software may implement the same function differently and leave different unique and detectable artifacts; The sequences of manipulations are not necessarily interchangeable; Anti-forensics applied during the manipulation obfuscates detection indicators; detection algorithms target specific types of manipulations and residual artifacts including light changes (artificial sources), semantic discrepancies (e.g. the Eiffel Tower in New York City), and compression effects on distributions within the Fourier domain. (3) The evaluation ground-truth data differs from historical data provided by some media editing programs, such as Adobe Photoshop or GIMP's history log files, layers, or other event recording scripts in the following aspects: The data required by evaluation differs greatly from that of software logs; The log file is an incomplete representation of all manipulation steps. It misses key data items and does not delineate backtracked or undone work; some evaluation sensitive data is not recoverable from log files or software image files; The log file is loose structure presentation, not suited for media forensic evaluation purposes; Log and script files detail software specific operation names not representative of all possible operations across the growing set of manipulation software and algorithms. Thus, these software suites do not sufficiently generalize manipulation algorithms necessary for evaluation factorization.

In this paper, we present an extensive three-year design and development project to create a novel media manipulation journaling tool that automatically or semi-automatically can collect, generate, and annotate the data and detailed manipulation metadata. JT uses a graph representation to support: i) recording the manipulation history; ii) integrating data collection, annotation, and its metadata collection, and iii) generating automated media manipulations using batch processes with pre-established manipulation sequences.

2. Evaluation data collection requirements

In addition to the labor intensity of media manipulation and metadata collection and labeling, JT is designed to support a diverse array of evaluation tasks, such as manipulation detection, localization (providing the localized manipulated region of the media), and manipulation history graph reconstruction.

Each task requires a wide variety of data and metadata to support performance evaluation and analysis using a multi-factor analysis approach. The primary requirements for image manipulation detection and localization task are: Manipulation history including intermediate images, manipulation software, operations, and its metadata; Origination data including camera, lens, environment, collection time, location etc. Semantic annotation and meaning to capture the purpose of a manipulation or series of manipulations designed to achieve a specific goal. Annotations include data identifying subject matter and setting of media. Semantic metadata includes events, weather, seasons, and intended effects of manipulation such as adding shadow or lighting inconsistency; Re-

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

compression including camera emulation as well as simulated and real-world social media images given specific procedures (e.g. Facebook image upload or download) to create realistic testing data for applications; Dynamically generated reference ground-truth mask for manipulated region.

3. Journaling Tool

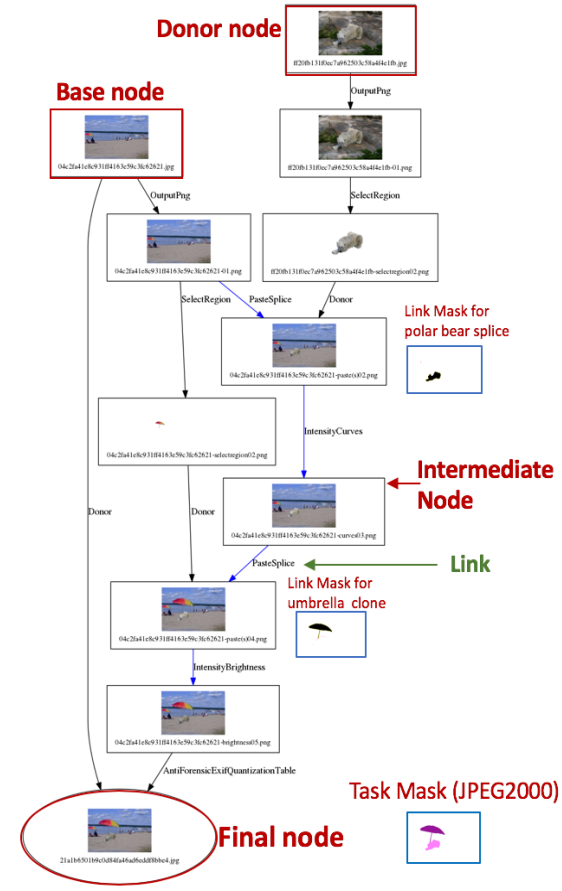
The Journaling Tool is a unified framework for data and metadata collection, annotation, and generation of automated manipulations designed according to data collection requirements. The intent of journaling is to capture a detailed history graph for each media manipulated project that results in a set of one or more final manipulated media files. The data collection process requires media manipulators to capture the detailed steps of manipulations during the manipulation process. In order to reduce the burden on manipulators, automation is built into the capture process to record incremental changes via mask generation and change analysis.

We designed a data collection approach to represent the media manipulation history with a Directed Acyclic Graph (DAG) to store manipulation history and metadata with the aim to maximize both human and machine intelligence effectively to perform multiple types of data collection and annotations. The data and metadata are collected in a hierarchical structure with three levels including a journal level, link level (a serial of operations for a given manipulated image), and node (image) level. Graph analysis algorithms support applying transformation rules to realign every image in the path from the original image to the final manipulated image, and back. The realignment provides a mapping of the original manipulation's downstream effect to the final media. The DAG structure supports reference mask generation based on different evaluation criteria.

The JT framework supports data collection and annotation throughout all stages of the manipulation process. Before manipulation begins, task design and media information are collected. During manipulation, details of each operation are captured. And upon completion post-processing produces target masks aligning operation changes to the final media product. This framework forms a comprehensive collective product we call a 'project'. During the manipulation process, details of each manipulation operation is captured that would otherwise be lost after manipulation is completed; Upon completion, post-processing produces target masks aligning operation changes to the final media products.

3.1. Manipulation history representation using DAG

In a DAG produced by JT, a node represents a media file instance such as an image, video, or audio file. An edge, referred to as a link in this paper, represents an operation that altered the source node's media to produce the destination node's media. In the general sense, the link represents a function that consumes the source and produces the destination. All metadata associated with the function is maintained with the link, including additional parameters, semantic information and change analysis. However, it is more accurate to generalize the link as a dependency between source and destination, such that the destination depends directly on the state of the source. The DAG forms a dependency tree, and, by nature of its construction, records the sequence of operations used to produce manipulated media from non-manipulated media.



Base node image Final node image

Figure 1: An example of a human journal (All images, graphs, and charts are original works created under contract on the MediFor Program [9]).

Figure 1 shows an example of a human manipulated journal. There are four types of nodes: base, donor, final and intermediate. A base node represents the primary media

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

(original) being altered whereas the donor represents media contributing the alteration of the base. Base and donor nodes do not have predecessors. Base nodes represent non-manipulated, "high-provenance", media that is camera original without any processing after capture. Donor nodes are images with un-specified provenance. Final nodes do not have successors; each final node represents a final product of a sequence of manipulations. All other interim nodes document the state of the media produced by a single manipulation.

Links form the dependencies between each manipulation state of the media. There are two kinds of links: operation and donor. An operation link represents an operation performed on a source node media file to produce a manipulated result. A donor link represents the donation of one media to the alteration of another, such as a 'paste' type operation would require. Although the donor is conceptually a parameter to the 'paste' operation, the link forms the necessary dependency.

The tool enables manipulators to record intermediate states of the media during the manipulation process, recording the incremental changes from each state to the next. Steps may be grouped to align to a semantic purpose, such as steganography.

Incremental changes include differences measured for each pixel, variations over time (video and audio), metadata alterations, and data in support of realignment of a manipulation mask to the final media including affine transform reconstruction. For example, filling a region within an image prior to a resize is portrayed as the proportionate region in the resized image.

A basic manipulation unit performs one cohesive function on the media. The operation is recorded with the name and version of the software used to perform the operation, and the parameters used in the operation. Upon completion of the operation, a mask is generated to record the specific changes made by the operation. This serves to identify the affected data of the operation and validate the operation, identifying accidental side effects. Operations are grouped into categories for ease of identification by manipulators. Paste-Splice operations involve the donation of pixels from another image. To facilitate easy recognition of the donated pixels, manipulators perform a select-region operation, applying an alpha-channel to only select donated pixels, prior to paste.

Journals are organized into patterns. These patterns set standards to avoid easily detectable side-effects. For example, images are first converted to a lossless RGB format such as Portable Network Graphics (PNG), manipulated, and then converted back to the desired media type, often applying anti-forensic operations such as re-applying specific camera quantization tables and EXIF alterations, cleansing telltale signs of manipulation.

3.2. Masks

Given a manipulated test media, its reference ground truth mask for the manipulation localization task is needed in the evaluation. The reference mask is an image where each pixel indicates whether the associated pixel in the test media has been manipulated or not. If the media was manipulated by a series of manipulations, then the reference mask is a composite mask which aggregates all manipulations' masks along the path from a base image to the given test media in the final node. The composite mask has the same dimension with the test image and is represented in JPEG2000 format. Each channel of the JPEG2000 represents a manipulation operation mask aligned with the test image. The reference mask is aligned to the test media for uniformity over all operations including seam carving and cropping, for which the mask describes pixels removed.

In order to obtain the reference mask, up to four types of link level masks are generated: (i) Input Mask: provided by the manipulator as metadata, the mask is composed of the alpha channel of the portion of the image changed or selected, depending on the operation. The input mask is interpreted based on the operation. For Paste Clone operations, the input mask reflects the selected cloned pixels. For seam carving, the input mask reflects the protected pixels, which the manipulator does not want to change. In this case, the relative position and intensity of each pixel does not change with respect to other pixels identified in the mask.

(ii) Difference Mask: indicates the differences between the before and after manipulation operation. It was generated by capturing pixels changed during the manipulation. For the Crop operation, the difference mask reflects the change in the cropped pixels, which is expected to be none. For seam carving, the difference mask reflects the removed seams. Since full reconstruction of removed seams along two dimensions is difficult, the JT is equipped with a seam carving algorithm that records the specific seams.

(iii) Task Mask: a task-specific mask identifies the affected pixels of a final test image for a given link operation (not all link operations contribute to a task-specific mask). Global operations, such as blur and transforms, are excluded from task-specific masks. The construction of the task-specific mask includes application of all subsequent transforms to a link's difference mask up to the final image node including, but not limited to: Resize, Rotate, Warp, Affine, Crop, Flip, Cut/Remove/Carve, and Content Aware Scale operations. In the case of seam carving where the removed seams are determined, the task mask represents those pixels neighboring removed pixels along seams.

(iv) Donor Mask: A task specific mask that identifies the donated pixels from a non-manipulated donor image. As transformations may occur prior to donation, the base image donor mask is constructed by applying antecedent transforms to the donor link's difference mask.

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

The difference mask for a donor link reflects the set of pixels from the donor image pasted into an image. During the paste operation, the pasted image may be cropped, rotated and resized. Thus, the donor mask may not necessarily reflect the selected region prior to paste splice. Often the selected region from donor pixels does not represent exact pixels donated in a paste splice. In these cases, SIFT/RANSAC operation is used to determine a perspective transformation applied to the paste splice mask to produce an accurate donor mask.

3.3. JT Algorithms

JT combines human and machine intelligence to optimize the collection process. Several automatic algorithms have been developed to reduce the need for human annotation. ***Link level mask analysis***: each operation triggers operation specific analysis. All operations are concluded with structure similarity, peak signal to noise ratio and categorizations of size of change (small, medium, large). Transforms include SIFT/RANSAC computations to construct a perspective transformation matrix. Resize records the size change. Crop records both the size change and the location of the upper left pixel of the cropped area from the source image. ***Automatic single operation mask generation***: for example, local operation mask (e.g. Fill), splice paste mask, splice donor mask, and seam carving mask. ***Automatic target mask generation***: Mask generation based on the specified subtask. ***Automatic metadata generation***: journal level metadata (e.g. link count, journal complexity) and image level metadata (e.g. manipulation unit count, image complexity, manipulation summary, manipulation size).

3.4. BatchJT, AutoJT, and ExtendedJT

Evaluating specific capabilities of each detector and supporting the training of machine learning based detectors requires a good distribution and a variety of factors. Capturing detail-rich on-the-fly manipulation data adds additional burden to the manipulator, impeding the speed of producing manipulated media products.

The JT embodies three core components to automate manipulations either from start to finish or by extension of human manipulated products to quickly expand the breadth of a dataset. The first is a pipeline-based batch tool (BatchJT) to automate the creation of journals in accordance to a graph specification. The second is an automatic graph specification generating tool (AutoJT) that generates permutations of graph specifications for production of many controlled variations of journals. The third is the extended journaling tool (ExtendedJT) to automate the extension of a manipulation graph producing additional branches of manipulations off selected nodes with a set of scripted operations.

Figure 2 shows an example of an AutoJT journal (randomly generated graph). AutoJT can mimic human-based journals to produce a large number of diverse manipulation data to support statistical analysis using a wide range of manipulation types with different parameter values.

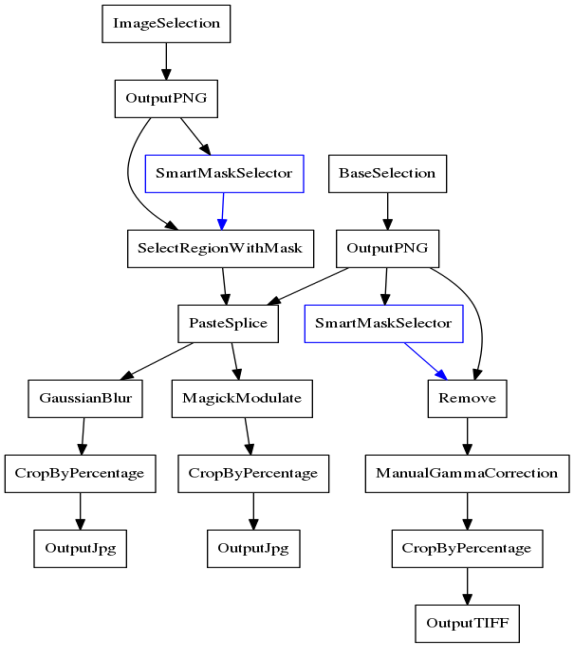


Figure 2: An example of an AutoJT journal graph

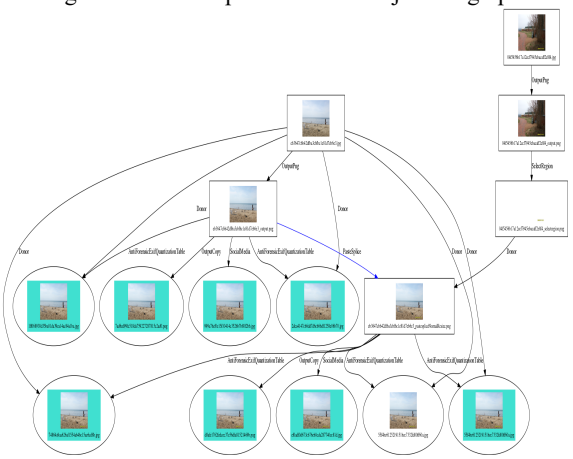


Figure 3: An example of an ExtendedJT journal graph.

Figure 3 shows an example of an ExtendedJT journal graph that extended a human-generated journal with four operations for each intermediate node (Facebook laundering and anti-forensics) with two different parameters, and saved in PNG format.

Through automation, permutation over the factor characteristics provide wider coverage of variability, such as evaluating crop detectors, where images are cropped with varying sizes and positions.

Furthermore, as many media transforms may be

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

automated, an automation framework assisted by DAG specifications can direct manipulation plugins, sequenced together in controlled combinations to produce both the manipulated media and the accompanying DAGs in large batches.

3.5. JT key features

The JT generalizes mostly commonly used manipulation operations (for image, video, and audio), aligned to categories for consistent meaning across both manipulation and detection points of view.

The data and metadata collected by JT is classified into the following categories: (i) Media and supporting metadata, camera model, location, time, etc. (ii) Annotations providing semantics of individual or a sequence of manipulations; (iii) Time-of-recording manipulation masks and associated analysis data.

Metadata is organized and collected into three levels: link level, final node level, journal level. Link level captures the specific operations and change analysis of the affected medium. Final nodes capture summary information of all operations applied to a media leading to its final state. Journal properties capture the overall intent of the journal including semantics and complexity.

JT has the following major functions:

- Collect manipulation history data into a DAG;
- Generate link level evaluation masks given different types of manipulation operations;
- Generate the final image composite evaluation mask;
- Automatically generate manipulation journals given the manipulation operations and DAG graph with resources (the base image, manipulation operation and its parameters, etc.);
- Auto-Extension of existing journals;
- GAN [22] image and video journaling with GAN tools;
- Automatic video temporal frame-drops journaling for frame drop operation;
- A notification system to integrate with task management services such as project management type services and email systems;
- Validation components for quality control;
- Integrated manipulation detection tools for manipulators to assess the quality and detectability of their manipulations;
- A rich plugin architecture for adding operations, media readers (e.g. raw formats), validation rules, manipulation detection tools and remote notification systems.

3.6. JT advantages

As human manipulation on images and videos is costly and time consuming, where possible, JT reduces the burden of annotation without compromising the fidelity of the historical data. JT is a unified framework that guides the journaling process to ensure consistent and quality journals through employing the following seven concepts: (i) Concise operation definitions for all manipulation operations along with required parameters and allowed responses. (ii) Validation rules to capture mistakes in the journaling process (e.g. resize during a format change). (iii) Quality assessment tools to ensure that donor and target masks can be aligned to donor and final image nodes, respectively, given the recorded transforms. (iv) Application of anti-forensics along with effectiveness measures to support a quantitative measure of manipulation detection difficulty. (v) ExtendedJT to quickly expand the test dataset, which uses all intermediate images generated and collected to serve as probes for different evaluation tasks. For example, one journal may contain more than 50 intermediate images associated with one final manipulated image representing the sequence of operations including blur, splice in-painting, and other transforms. Those intermediate images serve for evaluation on specific operations and the combination of those operations (e.g. splice followed by remove). (vi) Facilitate reuse and expansion of journals through extensions applied to intermediate node as required by the evaluation and/or training tasks. (vii) Automatically generate journals with a designed graph structure.

JT is publicly available as an open source package maintained on github (https://github.com/PAR-Government/media-journaling-tool). It is implemented in Python and has a detailed user guide.

4. Evaluation

4.1. Dataset generation tool: TestMaker

Given all the resources that the data collection team collected, manipulated, and annotated, the next step is to build the evaluation datasets for the task evaluations. (Please refer to [10] for all task definitions.) TestMaker is a tool to generate the evaluation test datasets with reference ground-truth data defined in [22] and used for evaluation scoring packages, MediScore, (https://github.com/usnistgov/MediScore). At the same time, TestMaker also validates journals (quality control) and metadata produced by JT and construct evaluation dataset. We will describe TestMaker in details in another document.

One of the evaluation requirements is called “selective scoring”; that is, to select a subset of data defined by a query condition (target manipulations) from the whole test pool to score a system. For the example, in Figure 1, if one would like to evaluate the performance of a Copy-Move detection system on copy-clone only images, the final image is selected as the test image, and the reference ground-truth region for evaluation is only the umbrella region.

4.2. Preliminary experiment on mask collection

As discussed in Section 1.2, post annotation of manipulated media does not capture sufficient detail to meet the needs of the evaluation program. Figure 4 demonstrates why it is important to collect and verify the mask during the manipulation process and also why most post-annotation image editing software could not provide the correct mask used for the evaluation. The first row is the

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
”Manipulation Data Collection and Annotation Tool for Media Forensics.”
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

original image (from the base node), the donor image (from the donor node) with a yellow ball, and the final manipulated image which pastes (or splices) the yellow ball into the first image (from the final node).

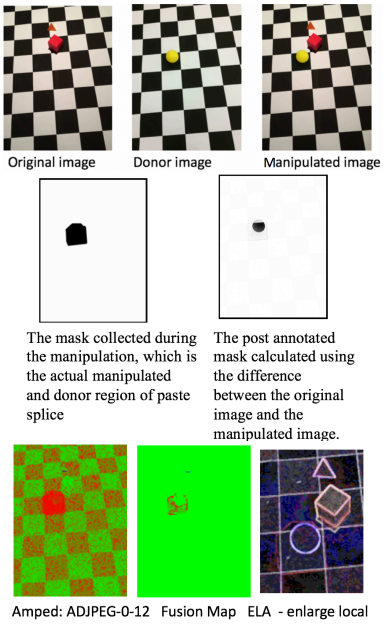


Figure 4: Mask collected during manipulation vs. generated after manipulation (post-annotation).

Notice that the imaging conditions of the base and donor images are very similar. The manipulator cut a polygon region from the donor image with the ball and directly pasted it to the base image. The modified pixel region should be the polygon region as shown in the first image of the second row. After the manipulation is done, there are three images: the base, donor, and final manipulated images. Using the images, one could perform post annotation to generate the mask of the manipulated region by calculating the difference of the base and manipulated images, which is the second mask of the second row with a half ball in it. This manipulation mask reflects neither the manipulated region nor its boundaries as detected using traditional detection algorithms (three algorithms in Amped Authenticate Software) as shown in the third row shows that the detection results for both region-based approach (ADJPEG) or boundary-based approach (ELA) are consistent with the mask collected during the manipulation, and are not consistent with the mask generated by post annotation. The locally enlarged ELA algorithm result shows the manipulation boundary with a red border, but the post-calculated mask cannot reflect this and so should not be used as ground-truth in the evaluation. Furthermore, such information is not captured in any image editing software and their history logs, such as Adobe Photoshop PSD files and logs. JT is designed to collect the evaluation ground-truth as required by the evaluation tasks.

4.3. Reference ground-truth masks for selective scoring

For test construction, the JT aligns manipulation masks to the evaluation media. Given a test image, the evaluation task masks for each manipulation are condensed into JPEG2000 containers in which each link mask is a bit plane. JPEG2000 is an image coding system that offers an extremely high level of scalability and accessibility. The standard supports precisions as high as 38 bits/sample. In our design, JPEG2000 was adopted to record distinct manipulations at any level for each test image, with each bit representing a distinct manipulation (represented by a distinct color in reference mask). The metadata associated with the container describes the bit plane used for each manipulation. A manipulation mask may also be associated with more than one-bit plane if the mask traverses through different transforms for two or more final manipulated media. For example, a paste splice mask may be followed by a seam carve in one evaluation media and a warp in another.

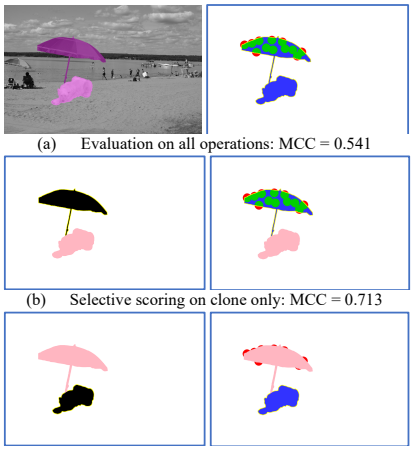


Figure 5: JPEG2000 mask for selective scoring evaluation

In the journal shown in Figure 1, the top beach image is the nonmanipulated base image. JT generated two local masks for the splice of the polar bear and the clone of the umbrella with their own bit plane value, expressed in the two individually colored masks in the left image of Figure 5 (a). A system output mask is shown in the right image. If the evaluation task is to detect all manipulated pixels regardless of manipulation type, then the ground-truth covers every manipulated region (all colors as shown in Figure 5 (a)). The Matthew Correlation Coefficient (MCC) of the system output mask is 0.541. If the evaluation task is to selectively evaluate only the clone detection system, then only the "clone" operation's mask should be used (the black region in the left of Figure 5 (b)) as ground-truth mask for the evaluation. The MCC of the same system output of the selective scoring on clone is 0.713. If the evaluation task is to selectively evaluate only the splice detection systems, then only "splice" operation's mask should be used (the

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

black region in the left of Figure 5 (c)) and the selective scoring result on splice is 0.

4.4. Manipulation history graph

JT provides the accurate ground truth phylogeny graph for the evaluation of provenance building systems—those systems retrieve related images with respect to a given query image from a world dataset and construct a phylogeny graph. The world data set is composed of random images downloaded from internet and the images from the journals. To measure the accuracy of the system output, it is compared to the reference ground-truth phylogeny graph derived from the JT journal graph. Figure 6 shows an example of the evaluation results for the history graph generation system. The green boxed images are correctly retrieved nodes, green links are correctly identified links, red boxed images are incorrectly retrieved nodes, grey boxed images are non-retrieved nodes, and grey links are missing links.

Figure 6: The evaluation result of a phylogeny graph produced by a provenance building system

4.5. Evaluation dataset summary

Table 1: A summary of released MediFor datasets

Image Dataset	Test Image # (K)	Journal #	Date
2017 Dev.	3.5	394	04/2017
2017 EvalPart 1	4	406	06/2017
2018 Dev1	5.6	178	12/2017
2018 Dev2	38	432	01/2018
2018 EvalPart1	17	758	03/2018
Video Dataset	**Test Video #**	**Journal #**	**Date**
2017 Dev.	214	23	04/2017
2017 EvalPart 1	360	47	06/2017
2018 Dev1	116	8	12/2017
2018 Dev2	231	36	01/2018
2018 EvalPart1	1036	114	03/2018

With JT, we have generated over 4500 human manipulated image journals and 400 video journals with different image and video editing programs with over 100 manipulation operations in diverse groups.

Using the manipulated image and video journals, we have generated several major evaluation datasets in the past two years. Table 1 summarizes released datasets. About 68K test images and 2K test videos with 2100 image journals and 200 video journals are available to the public. The table shows the public released dataset (one third of all evaluation data). We also have the corresponding sequestered datasets for sequester evaluation.

The metadata collected within a journal supports multi-dimensional system evaluation analysis. We can evaluate and analyze system performances by comparing across (1) different parameters for compression, image quality, resize, format, and image normalization (2) different manipulation types including splice, clone, remove and Content Aware Fill; (3) different manipulation software and algorithms including commercial off-the-shelf software, GAN, social media, etc. (4) different content type and presentations including faces, people, landscape, objects with different sizes, etc. (5) different manipulators with different skill levels and sets (6) different orders of manipulations and (7) different scanner, camera models, monitor, and printer medium, when considering recaptured media.

5. Discussion and future work

We are continuing to collect data and journals to support evaluations in future years. The design philosophy could be applied to other research domains. The JT packages are able to be adapted to other applications and purposes such as machine learning training data generation. We hope our JT framework will spur innovation in data collection and enable data-driven machine learning approaches applied to computer vision applications.

6. Acknowledgement

The authors gratefully acknowledge the members of the DARPA Media Forensics (MediFor) Program and members of the Air Force Research Lab for managing the program and weekly data team meetings; special thanks go to Matthew Turek, Neil Johnson, David Doermann, and Rajiv Jain for their instructions and strong support. The authors would like to thank Wendy Dinova-Wimmer for the initial experiments and design. PAR Government conducted this work under DARPA sponsorship via Air Force Research Laboratory (AFRL) contract FA8750-16-C-0168. NIST conducted this work under NIST Interagency Agreement Number 1505-774-08-000.

7. Disclaimer

Certain commercial equipment, instruments, software, or materials are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the equipment, instruments, software or materials are necessarily the best available for the purpose.

The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

References

[1] A. Piva, "An Overview on Image Forensics," *ISRN Signal Process.*, vol. 2013, pp. 1–22, 2013.

[2] H. Farid, "Photo forensics." The MIT Press, 2019.

[3] J. Fridrich, "Digital image forensics using sensor noise," *Signal Proc. Mag. IEEE*, vol. 26, no. 2, pp. 26–37, 2009.

[4] J. Fridrich, "Digital image forensics," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 26–37, Mar. 2009.

[5] M. Mishra and F. Adhikary, "Digital image tamper detection techniques-a comprehensive study," *ArXiv Prepr. ArXiv13066737*, 2013.

[6] H. Farid, "Image forgery detection," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 16–25, Mar. 2009.

[7] M. C. Stamm, Min Wu, and K. J. R. Liu, "Information Forensics: An Overview of the First Decade," *IEEE Access*, vol. 1, pp. 167–200, 2013.

[8] H. T. Sencar and N. Memon, "Overview of state-of-the-art in digital image forensics," *Algorithms Archit. Inf. Syst. Secur.*, vol. 3, pp. 325–348, 2008.

[9] M. Turek, N. Johnson etc., DARPA Media Forensics (MediFor) Program, https://www.darpa.mil/program/media-forensics.

[10] H. Guan; M. Kozak; E. Robertson; Y. Lee; A. N. Yates; A. Delgado; D. Zhou; T. Kheyrkhah; J. Smith; J. Fiscus, "MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation," in 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pg. 63-72, 2019.

[11] C. Halaschek-Wiener, J. Golbeck, A. Schain, M. Grove, B. Parsia, and J. Hendler, "Photostuff-an image annotation tool for the semantic web," in Proceedings of the 4th international semantic web conference, 2005.

[12] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," Int. J. Comput. Vis., vol. 77, no. 1–3, pp. 157–173, May 2008.

[13] C. Vondrick, D. Patterson, and D. Ramanan, "Efficiently scaling up crowdsourced video annotation," International Journal of Computer Vision, vol. 101, pp. 184-204, 2013.

[14] B. Sekachev, M. Nikita, Z. Andrey et al., Computer Vision Annotation Tool (CVAT), [Online] Available: https://github.com/opencv/cvat, description link: Computer Vision Annotation Tool: A Universal Approach to Data Annotation, https://software.intel.com/en-us/articles/computer-vision-annotation-tool-a-universal-approach-to-data-annotation, March 1, 2019, [Accessed: 08-Apr-2019].

[15] A. Dutta, A. Gupta, and A. Zissermann, "VGG Image Annotator (VIA)," 2016. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/software/via/. [Accessed: 08-Apr-2019].

[16] UC Berkeley DeepDrive project (https://deepdrive.berkeley.edu), "Scalabel (https://www.scalabel.ai)," 2018, [Online] Available: https://github.com/ucbdrive/scalabel. [Accessed: 08-Apr-2019].

[17] Anting Shen, "BeaverDam: Video Annotation Tool for Computer Vision Training Labels", EECS Department, University of California, Berkeley, Master Thesis, Dec. 2016. [Online] Available: https://github.com/antingshen/BeaverDam, [Accessed: 08-Apr-2019].

[18] D. Acuna, H. Ling, A. Kar, S. Fidler, "Efficient Interactive Annotation of Segmentation Datasets with Polygon-RNN++", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 859 – 868, 2018.

[19] Microsoft, "VoTT: Visual Object Tagging Tool", [Online] Available: https://github.com/Microsoft/VoTT, [Accessed: 08-Apr-2019].

[20] Google, "Cloud Video Intelligence - Video Content Analysis," Google Cloud Platform. [Online]. Available: https://cloud.google.com/video-intelligence/. [Accessed: 29-Mar-2019]

[21] P. Rennert, O. M. Aodha, M. Piper, and G. Brostow, "VideoTagger: User-Friendly Software for Annotating Video Experiments of Any Duration," Cold Spring Harbor Laboratory, 2018.

[22] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," International Conference on Learning Representations, 2018.

[23] A. Yates; H. Guan; Y. Lee, A. Delgado, D. Zhou, J. Fiscus, Media Forensics Challenge 2018 Evaluation Plan, NIST, 2018.

Robertson, Eric; Guan, Haiying; Kozak, Mark; Lee, Yooyoung; Yates, Amy; Delgado, Andrew; Zhou, Daniel; Kheyrkhah, Timothée; Smith, Jeff; Fiscus, Jonathan.
"Manipulation Data Collection and Annotation Tool for Media Forensics."
Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. June 15, 2019 - June 20, 2019.

SIS Contagion Avoidance on a Network Growing by Preferential Attachment

Brian Cloteaux
brian.cloteaux@nist.gov
National Institute of Standards and Techonology
Gaithersburg, Maryland

Vladimir Marbukh
vladimir.marbukh@nist.gov
National Institute of Standards and Techonology
Gaithersburg, Maryland

ABSTRACT

The economic and convenience benefits of interconnectivity drive the current explosive growth in networked systems. However, as recent catastrophic contagious failures in numerous large-scale networked infrastructures have demonstrated, interconnectivity is also inherently associated with various risks, including the risk of undesirable contagions. This paper reports on a work-in-progress on network formation subject to Susceptible-Infected-Susceptible (SIS) contagion risk mitigation. As opposed to existing research, we concentrate on network growth. Using a generalized form of preferential attachment, we consider evolving networks where each incoming node to the network combines a preference for connecting to high degree nodes with an aversion for contagion risk. Our initial simulation results indicate that contagion risk aversion significantly alters the topology and contagion propagation for an emerging network. We also discuss the computational aspects of our simulation and our future plans to extend this model.

CCS CONCEPTS

• **topology analysis and generation**; • **networks → network reliability**;

KEYWORDS

network-growth; preferential attachment; SIS contagion risk avoidance

ACM Reference Format:
Brian Cloteaux and Vladimir Marbukh. 2019. SIS Contagion Avoidance on a Network Growing by Preferential Attachment . In *2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) (GRADES-NDA'19), June 30-July 5 2019, Amsterdam, Netherlands.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3327964.3328502

1 INTRODUCTION

The economic and convenience benefits of interconnectivity drive the current explosive emergence and growth of networked systems. However, interconnectivity is inherently associated with various risks, including the risk of an undesirable contagion [5]. Network formation by strategic agents/nodes is affected by numerous competing incentives, including the incentives to directly connect to the "central" nodes and also to reduce risks of undesirable contagions. The effect of these two competing incentives on the network evolving by rewiring has been investigated in recent papers on adaptive and active networks [4, 8]. However, due to economic or contractual constraints, e.g., in the physical or financial networked infrastructure, breaking old and establishing new links may be difficult. In such cases, the effect of competing incentives on network growth plays a primary role in understanding of the structure and function of the emerging networked systems.

While in our future work we plan to investigate effect of competing incentives on networks simultaneously evolving by growing and rewiring, in this short paper we concentrate on growing networks subjected to an undesirable Susceptible-Infected-Susceptible (SIS) infection. We model multiple competing incentives by utility function and assume that network growth follows logit response dynamics [1]. This model describes bounded rational agents who attempt to maximize their utility with controlled level of rationality. This approach overcomes interrelated limitations of oversimplification and intractability of the game-theoretic approaches [6]. The tractability of this approach is due to the fact that logit response dynamics take the form of generalized preferential attachment [7], which allows one to leverage an extensive body of results on preferential attachment models for growing networks. Note that the same is true for rewriting models for networks of fixed size.

This paper is organized as follows. Section 2 proposes a model for network formation by contagion averse agents. Section 3 describes the simulation setup and computational challenges. Section 4 discusses the initial simulation results. Finally, Section 5 briefly summarizes and outlines directions for future research.

2 MODEL AND PROBLEM STATEMENT

Consider a growing network, where nodes arriving at discrete moments $t = 1, 2, \ldots$ make decisions to connect to an existing node i with degree d in order to attempt to maximize the utility function

$$u_d(\delta_i) = (1 - h\delta_i) \ln \varphi_d, \tag{1}$$

where $\delta_i = 1$ if the node i is infected and $\delta_i = 0$ otherwise, and where $\varphi_d \geq 1$ is an increasing function of $d = 1, 2, \ldots$. The parameter $h \geq 0$ characterizes the trade-off between the incentives for connectivity and for contagion risk avoidance. If $h > 1$, then the decision of an arriving node to connect to an existing infected node of degree d results in not only the loss of the utility of connectivity, but also in an additional loss $(h-1) \ln \varphi_d$. Thus connecting to a high degree node has a high-risk but also a high-potential for reward to the arriving node.

Following the conventional approach, we model bounded rationality by assuming that an arriving node immediately develops a

GRADES-NDA'19, June 30-July 5 2019, Amsterdam, Netherlands
2019. ACM ISBN 978-1-4503-6789-9/19/06.
https://doi.org/10.1145/3327964.3328502

fixed number $m \geq 1$ of connections to a subset of m existing nodes. All the nodes in this subset are selected independently from each other using logit probabilities [1]

$$g_i \sim \exp[T^{-1} u_{d_i}(\delta_i)], \tag{2}$$

where d_i is the degree of i, and the "temperature" $T > 0$ characterizes the agents' rationality. A value $T \to 0$ corresponds to complete rationality, while a value $T \to \infty$ corresponds to complete randomness. For the utility (1), the logit probabilities are:

$$g_i \sim \left(\varphi_{d_i}\right)^{\frac{1-h\delta_i}{T}}. \tag{3}$$

Further in the paper we consider a specific function, $\varphi_d = d^\gamma$, where $\gamma > 0$, when attachment probabilities (3) take the following form:

$$g_i \sim d_i^{(1-h\delta_i)\alpha}, \tag{4}$$

where $\alpha = \gamma/T$. Since we are interested in the effect of infection averseness on network formation, we contrast two extreme cases: the case of infection risk indifference for the agents/nodes ($h = 0$), and the case of a very high infection risk aversion ($h \gg 1$).

From Krapivsky, Redner, and Leyvraz [7], if $h = 0$ then the resulting structure of a growing network depends on whether the parameter α is smaller than, greater than, or equal to unity. If $\alpha < 1$, this mechanism results in a stretched exponential node degree distribution while when $\alpha > 1$, then one "central node" is connected to nearly every other node in the network. When $\alpha > 2$, this "winner takes all" phenomenon is so extreme that the number of connections between "non-central" nodes is finite even in an infinite network. Finally, if $\alpha = 1$, then a power-law node degree distribution, $N_d \sim d^{-\nu}$, emerges. In this case, the parameter ν, where $2 < \nu < \infty$, controls the finer details of attachment probability.

A fundamental and still to a large degree open question is why numerous existing networked infrastructures are characterized by power law node degree distribution. In the context of the above generalized preferential attachment mechanism, the question is identifying additional mechanisms responsible for the tuning parameter ν, ensuring the specific value of the preference parameter $\alpha = 1$. We suggest that one of such mechanisms may be contagion risk avoidance by systemic risk averse strategic agents. Indeed, node infection probability among other things depends on node degree, and thus contagion risk avoidance affects the attachment probability.

We simulate network growing by generalized preferential attachment and subjected to SIS infection. Once node n becomes infected, it spreads infection to each of its neighboring nodes j at fixed rate $\lambda > 0$. Node recovery time is distributed exponentially with the same recovery rate μ for all nodes. We define the effective infection rate ρ as the ratio between the infection and recovery rates,

$$\rho \stackrel{def}{=} \frac{\lambda}{\mu}. \tag{5}$$

Alós-Ferrer and Netzer [1] showed that an SIS model is infection free if and only if

$$\rho \leq \frac{1}{\Gamma}, \tag{6}$$

where Γ is the Perron-Frobenius (P-F) eigenvalue of the adjacency matrix $A = \left(X_{ij}\right)_{i,j=1}^{N}$, where $X_{ij} = 1$ if nodes i and j are connected by an edge, and $X_{ij} = 0$ otherwise. Thus, the value $\rho = 1/\Gamma$ defines the threshold boundary for the infection-free region.

Ferreira, Castellano, and Pastor-Satorras [2] showed that an SIS contagion results in higher infection probabilities for higher degree nodes, at least in a case of random uncorrelated network. Thus, one may expect that SIS contagion aversion of strategic agents forming network growing by generalized preferential attachment *suppresses* the "winner takes all" phenomenon for $\alpha > 1$. Our goal in this paper is to confirm and quantify this conjecture by simulations.

3 SIMULATION SETUP AND CHALLENGES

In order to examine the effects of SIS contagion aversion, we simulate the growth of a number of networks across varying values of α and ρ. For our simulations, we fix the recovery rate at $\mu = 0.1$ and vary the infection rate λ by varying ρ.

We examine two cases: one where we avoid connecting to any node that is currently infected, in other words, preferential attachment with infection avoidance (IA) (where the value $h \gg 1$ for equation (4)). For the second case, we ignore the infection state of a node when choosing where to connect ($h = 0$ for equation (4)). This case is simple generalized preferential attachment (GPA).

In constructing the graphs used in our simulations, in each case we began with a path graph of 5 nodes. Nodes were then sequentially added to the graph using the generalized form of preferential attachment. In our simulations, we fixed the number of attached edges to $m = 5$. The associated α for each simulation run is reflected in the attachment weights ω for selecting edges as a new node is introduced.

The probability that an introduced node forms an edge with an existing node i is proportional to its weight ω_i. The weight ω_i for the node i is given by

$$\omega_i = \begin{cases} d_i^\alpha & \text{for pure preferential attachment (GPA)} \\ (1-\delta_i) \cdot d_i^\alpha & \text{for infection avoidance (IA).} \end{cases} \tag{7}$$

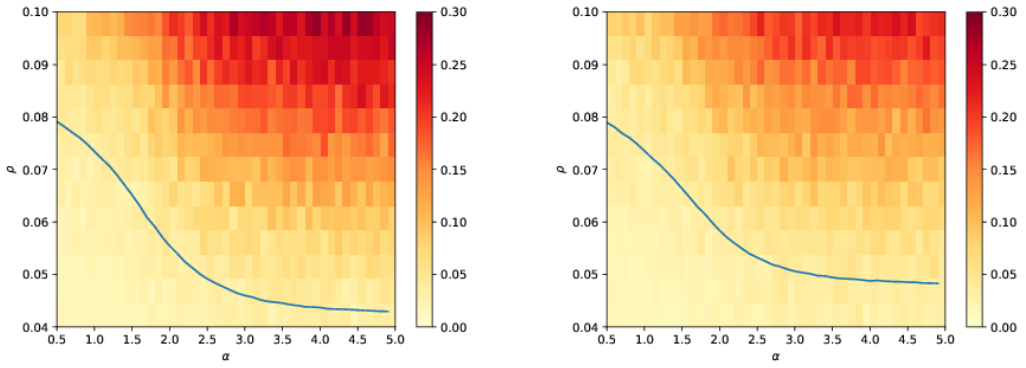
In order to attach the m edges of a new node, we select m unique nodes using these weights to define the probability distribution.

Between adding new nodes, we allow the SIS infection to reach a steady-state. In order to try to avoid correlation between infected nodes when we add new nodes, we allow the infection to run for 20 rounds after each node addition. Our analysis indicated that this value allowed the infection to reach a steady state before adding a new node.

For simulation runs where the value of ρ falls into the infection-free zone of (6), we risk the infection dying out after each SIS infection round. In order to prevent our infection avoidance simulations from degenerating into the simple case of preferential attachment, we maintain the infection state of the previous round. If at any time step the infection dies out of the network, we then simply backtrack to the last time step and rerun the simulation step to assure that the infection continues.

Each data point in our simulations is the average of 100 samples for a 100 node networks with the given α and ρ. Our simulation code is written in C++ using GSL [3] for the numerical computation library.

Cloteaux, Brian; Marbukh, Vladimir.
"SIS Contagion Avoidance on a Network Growing by Preferential Attachment."
Paper presented at 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) 2019, Amsterdam, Netherlands. June 30, 2019 - June 30, 2019.



(a) Without infection avoidance (GPA) **(b) With infection avoidance (IA)**

Figure 1: This figure shows the average portion of infected nodes across a number of α and ρ values.

4 INITIAL SIMULATION RESULTS

Localization of the principal eigenvector controls the spread of infection in a network [9, 10]. A network with some level of localization tends to maintain an infection over a subset of the nodes. Networks created using GPA show localization for their hub nodes. We are interested in whether IA networks also show localization, and how IA affects the spread of infection in these networks.

As expected, infection avoidance has a profound effect on the topology and infection sustainability for an evolving network. In Figure 1, we see the how infection is dampened when using infection avoidance. The intensity of the average infection is shown from 0 % to 30 % of the nodes for the ranges of α and ρ. Figure 1a shows simulation results while using general preferential attachment for growing the networks, while Figure 1b shows the results for networks grown using infection avoidance. The blue lines represents the boundary of the infection-free region given by $\rho = 1/\Gamma$. Each point in this line is computed by averaging the values for $1/\Gamma$ across all networks grown with a given α.

In comparing the upper right hand corners of the two graphs, the percentage of infected nodes in the resulting networks significantly decreases when using infection avoidance. We are also interested in the effect on the infection-free region, between using and not using infection avoidance growth. In comparing the two figures, we see that the threshold boundaries are raised in networks grown using infection avoidance. We note that the strongest effect on the resulting graphs occurs for α when $1 \leq \alpha \leq 2.5$. In the region $\alpha > 2.5$, localization begins to affect both regimes leveling out the threshold boundary. For the region $\alpha \leq 2.5$, infection avoidance slows the descent of the threshold boundary.

Since localization for the eigenvector of Γ depends on the maximum degree in a network, we plotted how the maximum degree is affected in Figure 2. In this figure, we show the ratio between the average of the maximum degrees from GPA networks over the IA networks. This ratio shows that in GPA graphs, the maximum degree grows much faster than in IA graphs when $\alpha \geq 1$. We also see that the strongest dependency for the maximum degree ratio

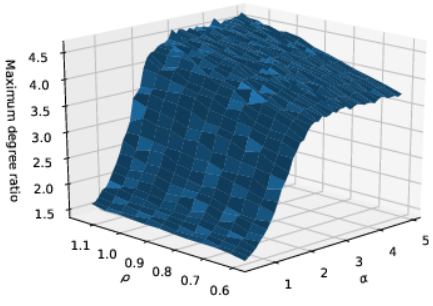


Figure 2: Ratio of maximum degrees between the GPA and IA graphs.

is on the value of α with a smaller effect from the value of ρ. Put together, these two figures imply that IA network growth limits contagion spread by limiting the amount of localization that occurs.

To test this idea we further examine the amount of localization that arises in IA networks by computing the inverse participation ratio (IPR) [9, 10]

$$IPR(\Gamma) = \sum_i \mathbf{f}_i^4(\Gamma), \tag{8}$$

where $\mathbf{f}(\Gamma)$ represents the eigenvector associated with the eigenvalue Γ. The IPR for a set of networks is usually studied as a function of the size of a network n by fitting it to a power-law distribution,

$$IPR(\Gamma) \sim n^{-\psi}. \tag{9}$$

Examining the slope ψ for a family of graphs helps us to understand the amount of localization occurring in these graphs. If the

Cloteaux, Brian; Marbukh, Vladimir.
"SIS Contagion Avoidance on a Network Growing by Preferential Attachment."
Paper presented at 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) 2019, Amsterdam, Netherlands. June 30, 2019 - June 30, 2019.

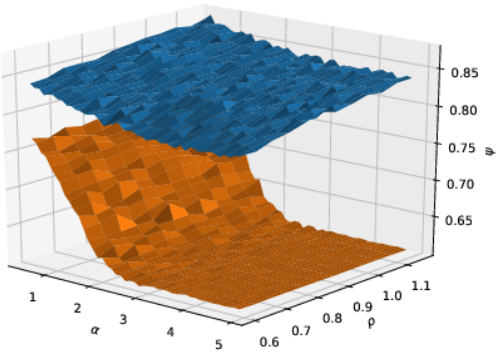


Figure 3: Slope ψ of inverse participation ratio between IA (blue surface) and GPA (orange surface) graphs.

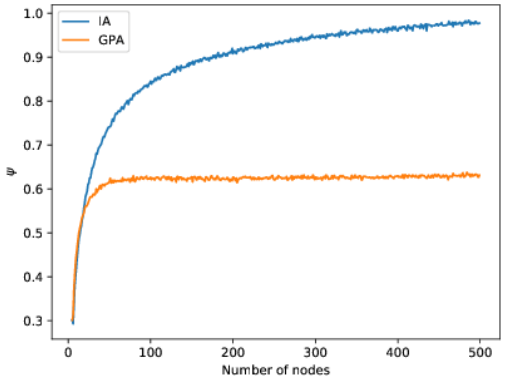


Figure 4: Slope ψ of the inverse participation ratio between IA and GPA graphs. Here the values $\alpha = 2$ and $\rho = 1$ are fixed.

slope ψ converges to 1, then no localization is occurring for the networks, else if the slope converges to value less than 1, then some localization occurs over a subset of the nodes.

In Figure 3, we see the that slope ψ is much higher for IA graphs than for GPA graphs. This figure shows that less localization is occurring in the IA graphs over all values of α and ρ. In Figure 4, we see a stronger indication that IA graphs do not show localization as they grow. Here we see that the value $\psi \longrightarrow 1$, implying that IA graphs do not demonstrate localization as the networks grow. Since $\psi \approx .6$ the GPA graphs do show some degree of localization.

5 CONCLUSION AND FUTURE RESEARCH

This paper reports on work-in-progress on the effect of SIS infection avoidance on a network growing by preferential attachment [8]. Our simulation-based investigation characterizes competing effects of the preferential attachment parameter α quantifying the propensity of an incoming node to connect to a high degree existing node on the one hand and effective SIS infection rate ρ on the other hand. Our initial simulation results suggest a possible explanation of the consistency of the structure of observed real networks with preferential attachment model.

The problem is that while typically observed networks are characterized by a power law node degree distribution, the preferential attachment model produces this node degree distribution only for the specific "preferential attachment parameter" $\alpha = 1$. For the example of SIS infection avoidance, we suggest a possible explanation for this phenomenon: even when the parameter $\alpha \neq 1$, preference for connectivity to high degree nodes combined with other incentives produces preferential attachment model with "effective preferential attachment parameter", $\alpha^{eff} \approx 1$. Indeed, infection avoidance counteracts the incentive to connect to a high degree node since high degree nodes are more likely to be infected.

Currently we are attempting to verify our conjecture that for $\alpha > 1$, infection avoidance keeps the network on the verge of eigenvector localization regime [9]. This conjecture may have important practical implications for infection propagation and mitigation on real networks since eigenvector localization implies persistent infection presence in network [10]. Our future plans include enhancing our model by allowing (a) connectivity decisions based not only on the current infected/non-infected node status, but also on the perceived likelihood of being infected, (b) rewiring of the already existing connections, (c) node investments in the infection risk mitigation, and (d) other types of contagion, e.g., threshold-based contagion.

REFERENCES

[1] Carlos Alós-Ferrer and Nick Netzer. 2010. The logit-response dynamics. *Games and Economic Behavior* 68, 2 (2010), 413–427. https://doi.org/10.1016/j.geb.2009.08.004

[2] Silvio C. Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. 2012. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Phys. Rev. E* 86 (Oct 2012), 041125. Issue 4. https://doi.org/10.1103/PhysRevE.86.041125

[3] Mark Galassi. 2009. *GNU Scientific Library Reference Manual - Third Edition* (3rd ed.). Network Theory Ltd.

[4] Thilo Gross, Carlos J. Dommar D'Lima, and Bernd Blasius. 2006. Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.* 96 (May 2006), 208701. Issue 20. https://doi.org/10.1103/PhysRevLett.96.208701

[5] Dirk Helbing. 2013. Globally networked risks and how to respond. *Nature* 497, 7447 (May 2013), 51–59. https://doi.org/10.1038/nature12047

[6] Matthew O. Jackson. 2008. *Social and Economic Networks.* Princeton University Press, Princeton, NJ, USA.

[7] P. L. Krapivsky, S. Redner, and F. Leyvraz. 2000. Connectivity of Growing Random Networks. *Phys. Rev. Lett.* 85 (Nov 2000), 4629–4632. Issue 21. https://doi.org/10.1103/PhysRevLett.85.4629

[8] Antoine Moinet, Romualdo Pastor-Satorras, and Alain Barrat. 2018. Effect of risk perception on epidemic spreading in temporal networks. *Phys. Rev. E* 97 (Jan 2018), 012313. Issue 1. https://doi.org/10.1103/PhysRevE.97.012313

[9] Romualdo Pastor-Satorras and Claudio Castellano. 2016. Distinct types of eigenvector localization in networks. *Scientific Reports* 6, 18847 (2016). https://doi.org/10.1038/srep18847

[10] Romualdo Pastor-Satorras and Claudio Castellano. 2018. Eigenvector Localization in Real Networks and Its Implications for Epidemic Spreading. *Journal of Statistical Physics* 173, 3 (01 Nov 2018), 1110–1123. https://doi.org/10.1007/s10955-018-1970-8

Cloteaux, Brian; Marbukh, Vladimir.
"SIS Contagion Avoidance on a Network Growing by Preferential Attachment."
Paper presented at 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) 2019, Amsterdam, Netherlands. June 30, 2019 - June 30, 2019.

CASFinder: Detecting Common Attack Surface

Mengyuan Zhang[1] , Yue Xin[1] , Lingyu Wang[1], Sushil Jajodia[2], and Anoop Singhal[3]

[1] Concordia Institute for Information Systems Engineering, Concordia University
wang@ciise.concordia.ca
[2] Center for Secure Information Systems, George Mason University
jajodia@gmu.edu
[3] Computer Security Division, National Institute of Standards and Technology
anoop.singhal@nist.gov

Abstract. Code reusing is a common practice in software development due to its various benefits. Such a practice, however, may also cause large scale security issues since one vulnerability may appear in many different software due to cloned code fragments. The well known concept of relying on software diversity for security may also be compromised since seemingly different software may in fact share vulnerable code fragments. Although there exist efforts on detecting cloned code fragments, there lack solutions for formally characterizing their specific impact on security. In this paper, we revisit the concept of software diversity from a security viewpoint. Specifically, we define the novel concept of *common attack surface* to model the relative degree to which a pair of software may be sharing potentially vulnerable code fragments. To implement the concept, we develop an automated tool, *CASFinder*, in order to efficiently identify common attack surface between any given pair of software with minimum human intervention. Finally, we conduct experiments by applying our tool to real world open source software applications. Our results demonstrate many seemingly unrelated software applications indeed share significant common attack surface.

1 Introduction

Code reusing is a common practice in today's software industry due to the fact that it may significantly accelerate the development process [6, 9]. However, such a practice also has the potential of leading to large scale security issues because a vulnerability may be shared by many different software applications due to the shared libraries or code fragments. A well known example is the *Heartbleed* vulnerability in *OpenSSL*, which caused widespread panic on the internet since the vulnerable library was shared by many popular Web browsers, including Apache and Nginx [10]. In addition to shared libraries, the reusing of existing code fragments may also lead to similar vulnerabilities shared by different software applications. Unlike libraries, such reused codes are typically not traced by any official documentation, which makes it more difficult to understand their security impact. Finally, this phenomenon may also compromise the well known concept of relying on software diversity for security, since seemingly unrelated software applications made by different vendors may in fact share common weaknesses.

⋆ This work was completed when the author was a Ph.D. student at Concordia University.
⋆⋆ This work was completed when the author was a M.Sc. student at Concordia University.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

The issue of identifying and characterizing the security impact of shared code fragments has received little attention (a more detailed review of the related work will be given in Section 6). Most existing vulnerability detection tools focus on identifying vulnerabilities for a specific software application based on static and/or dynamic analysis, with no indication whether different software may be sharing similar vulnerabilities due to common libraries or reused codes [11]. On the other hand, existing efforts on software clone detection mostly focus on identifying reused code fragments based on either the textual similarity or functional similarity, with no indication of the security impact [40]. Clearly, there exists a gap between the two, i.e., *how can we leverage existing efforts on software clone detection to characterize the likelihood that given software applications may share similar vulnerabilities?*

In this paper, we address the above issue through defining the novel concept of *common attack surface* and developing an automated tool, *CASFinder*, to calculate the common attack surface of given software applications. Specifically, we first extend the well known attack surface concept to model the relative degree to which a pair of software may be sharing potentially vulnerable code fragments. Such a formal model enables the quantification of software diversity from the security point of view, and its results may be used as inputs to higher level diversity methods (e.g., network diversity [47] and moving target defense [19]). Second, we develop *CASFinder* which is an automated tool that takes the source code of two software applications as the input and outputs their common attack surface result in an XML file or to a database. Third, we conduct experiments by applying our tool to a large number of real world open source software applications belonging to seven different categories from *Github*. More than 80,000 combinations of software applications are analyzed, and our results demonstrate many seemingly unrelated software applications indeed share a significant level of common attack surface. In summary, the contribution of this paper is threefold.

- First, to the best of our knowledge, this is the first effort on formally modeling the security impact of reused code fragments. The common attack surface model may serve as a foundation and provide quantitative inputs to higher level security-through-diversity methods.
- Second, the *CASFinder* tool makes it feasible to evaluate the common attack surface between open source software applications, which may have many practical use cases, e.g., providing useful references for security practitioners to choose the right combinations of software applications in order to maximize the overall software diversity in their networks, and reusing the knowledge about existing vulnerabilities in one software to potentially identify similar ones in other software.
- Third, our experimental results prove the possibility of similar vulnerabilities shared by seemingly unrelated software applications made by different vendors. We believe such a finding may help attract more interest to re-examining the concept of software diversity and its security implication.

The remainder of this paper is organized as follows. Section 2 provides a motivating example and background information. Section 3 defines the common attack surface model. Section 4 designs and implements the *CASFinder* tool. Section 5 evaluates the tool through experiments using real open source software. Section 6 reviews related work and Section 7 concludes the paper and provides future directions.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

2 PRELIMINARIES

In this section, we first present a motivating example in Section 2.1 and then provide background knowledge and highlight the challenge in Section 2.2.

2.1 Motivating Example

As an example, consider an enterprise network with Web servers running either the Apache HTTP server (*Apache*) or the Nginx HTTP server (*Nginx*), as well as a Cyrus IMAP server (*Cyrus*). Assume all three software applications are of the vulnerable versions that are affected by the *Heartbleed* vulnerability. This vulnerability has reportedly affected an estimated 24-55% of popular websites and gave attackers accesses to sensitive memory blocks on the affected servers, which potentially contain encryption keys, usernames, passwords, etc. [10]). The vulnerability is discovered inside the popular *OpenSSL* library, which is an extension of many Web and email server software applications for supporting the *https* connections.

Specifically, Figure 1 demonstrates how this vulnerability functions in relation to the three software applications in our example. Those software simply hand the encryption tasks to the *OpenSSL* extension, and the vulnerability appears when the software make external calls to the *OpenSSL* extension. In establishing the SSL connections, the API invocation *SSL_CTX_new(method)* is a function for establishing SSL content, *SSL_new()* is for creating SSL sessions, and *SSL_connect()* for launching SSL handshakes. To exploit the *Heartbleed* vulnerability, attackers would craft a heartbeat request with a special length and send it to the servers. This request would causes different software applications to invoke the same library function *memcpy()* without any boundary check enabling attackers to extract sensitive memory blocks from the servers.

The fact that this vulnerability exists inside the *OpenSSL* extension shared by all three software means an attacker can compromise those different software in a similar manner. This phenomenon is certainly not limited to this particular vulnerability. In this example, since both the *Apache* and *Nginx* projects are Web servers developed in C language, their similar functionality implies there is a high chance that the developers of both projects would not only import the same libraries, but also reuse the same or similar code fragments. In addition, as will be shown through our experimental results, code reusing also exists among software applications with very different functionalities. On the other hand, not all server software that use SSL connections are affected by this vulnerability, e.g., *Microsoft IIS* and *Jetty* are both immune to the vulnerability [10].

Clearly, there exists a need for identifying the software applications which may share such a common vulnerability, and for characterizing the level of such sharing since some software may share more than one such vulnerability. Such a desirable capability may have many practical use cases. For instance, it may allow similar software patches or fixes to be developed and applied to different software applications in order to mitigate a common vulnerability, which may significantly reduce the time and effort needed for developing such patches and fixes. This capability may also allow administrators to better judge the amount of software diversity in their networks, and to choose the right combinations of software applications (e.g., Apache and IIS w.r.t. this particular vulnerability) to increase the diversity. Finally, this capability would lead to a more

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

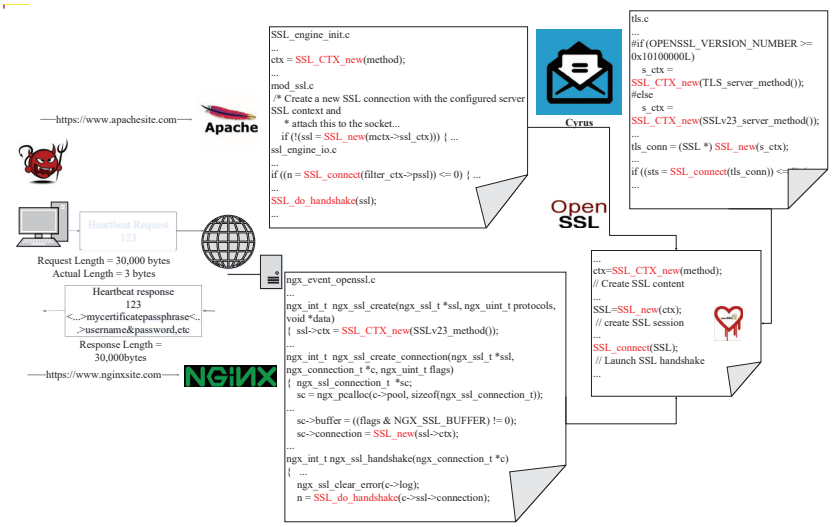


Fig. 1. An Example of the *Heartbleed* Vulnerability

refined approach to moving target defense (MTD) [12] since it could potentially allow us to quantify the amount of software diversity that is achieved by switching between different software resources under a MTD mechanism.

2.2 Background

We take two steps towards measuring the potential impact of cloned codes on security. The first step is to find similar code fragments in different software applications. The second step is to characterize the security impact of such code fragments. We first review some of the background concepts related to each step.

First, to detect similar code fragments between software, most clone detection methods are based on either the textual similarity or the functional similarity, and existing tools are mostly based on text, token, tree, graph, or metrics [40, 41]. Among the existing tools, we have chosen *CCFinder* [22], a language-based source code clone detection tool, to find cloned code fragments within given software. As one of the leading token-based detection tools, *CCFinder* has received the Clone Award in 2002, and it supports multiple languages, including C, C++, Java, and COBOL. *CCFinder* first divides the given source code into tokens using a lexical analyzer. It then normalizes some of those tokens by replacing identifiers, constants and other basic tokens with generic tokens representing their language role. Finally, it uses a suffix-tree based sub-string matching algorithm to find common subsequences corresponding to clone pairs and classes [22]. A key advantage of such a token-based tool is that it can tolerate minor code changes, such as formatting, spacing and renaming, in the reused code.

However, the result from clone detection tools, including *CCFinder*, only reveals similar code fragments between source codes, without indicating any security impact.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

The primary challenge is therefore to model and quantify the potential impact of clone detection on security in terms of leading to potential vulnerabilities. To this end, a promising solution is to apply the attack surface concept [35], which is a well known software security metric that measures the degree of software security exposure. The measurement is taken as counts along three dimensions, the entry and exit points (i.e., methods calling I/O functions), channels (e.g., TCP and UDP), and untrusted data items (e.g., registry entries or configuration files), and the counting results are then aggregated through weighted summation. Attack surface measures the intrinsic properties of a software application, e.g., how many times does each method invoke I/O functions (which provides an estimate of security risks such as buffer overflow), regardless of external factors such as the discovery of the vulnerability or the existence of exploit code. Therefore, attack surface can potentially cover both known and unknown vulnerabilities.

Therefore, we will combine clone detection (i.e., *CCFinder*) with attack surface to quantify the likelihood that cloned code fragments may lead to potentially similar vulnerabilities shared between different software applications. For simplicity, we will focus on entry and exit points in this paper, and will consider channels and untrusted data items in our future work. We also note that, since it is not guaranteed that every entry or exist point will map to a vulnerability, the attack surface concept is only intended as an estimation of the relative abundance of vulnerabilities in software [35]. Consequently, our model and tool also inherit this limitation, and the results will only indicate the potential, instead of the actual existence, of common vulnerabilities.

Combining the result of clone detection with the attack surface concept is not a straightforward task. We discuss a key challenge in the following. In Figure 2, function *handle_response()* and function *quicksand_mime()* are both entry points since they call I/O functions *fseek()* and *ftell* (from the standard C library). A naive application of the attack surface concept here would indicate each function count as one entry point and hence both have the same security implication. However, such a coarse-grained application ignores the exact number of I/O function calls (i.e., three calls in *handle_response()* and two in *quicksand_mime()*) whose difference may be significant in practice. In our model, we will take a more refined approach to address such issues.

```
1 fseek(fp, 0, SEEK_END);
2 size = ftell(fp);
3 fseek(fp, 0, SEEK_SET);
4 snprintf(fsize, 32, "Content−Length: %d\r\n\r\n", size);
```

```
1 long fsize = ftell(f);
2 fseek(f, 0, SEEK_SET);
3 free(decoded_mime);
```

Fig. 2. Examples of Entry Points: /Simple-Webserverche/server.c *handle_response()* (Top) and /quicksand_lite/libqs.c *quicksand_mime()* (Bottom)

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

3 The Model of Common Attack Surface

In this section, we model the security implication of cloned code fragments between software applications through two novel security metrics, namely, the *conditional common attack surface* (ccas) and the *probabilistic common attack surface* (pcas). Those two metrics are designed for different use cases as follows.

- The *conditional common attack surface* (ccas) is designed to be asymmetric for use cases in which one software is of particular interest and evaluated against all other software. For example, suppose a company has developed a new Web server application and wants to understand any similarity between their product and other existing Web servers such as *Apache* and *Nginx*. In such a case, the key is to rank those other software applications based on the relative percentage of shared attack surface, and the developer can apply the metric *ccas* for this purpose.
- Second, in a different scenario, suppose an administrator wants to understand the level of software diversity between all the software applications inside the same network. In such a case, both software in comparison are considered equally important, so the symmetric metric *pcas* would be more suitable, which will yield a unique measurement of shared attack surface between any pair of software. The following details the *ccas* and *pcas* metrics.

3.1 Conditional Common Attack Surface (CCAS) Metric

We first consider clone segments between two software applications identified using *CCFinder* [22] through an example.

Example 1. Figure 3 demonstrates clone segments between a Web server application *SimpleWebserver* and an ssh application *SSHBen*. In the figure, the *Clone id* is a unique number labelling a group of related clones inside both software applications. For instance, the code segments inside the solid line blocks indicate the clone segments with the same Clone id 28, and the dashed line blocks are for Clone id 78. Note that the same code may appear under different clone ids, e.g., line 146 and 147 in *Simple-Webserver* appear under both clone ids. Also note that, for Clone id 78, the matching between the two clone segments is *inexact* [22] since $strcat$ does not exist in *SSHBen*.

From the above example, it is clear that the clone segments belonging to the same Clone id are not identical between the two software applications. Therefore, the attack surface would be asymmetric as well. First, we define the *Common Attack Surface* as the collection of I/O function calls inside the clone segments as follows.

Definition 1 (Common Attack Surface). *Given two software applications A and B, the common attack surface of A w.r.t. B (or that of B w.r.t. A) under the Clone id i is defined as the multi-set (which preserves duplicates) of I/O function calls that exist inside the clone segments of A under the Clone id i, denoted as $cas_i(A|B)$ (or $cas_i(B|A)$).*

Example 2. To follow our example, we have

- $\text{cas}_{28}(\text{SimpleWebserver}|\text{SSHBen}) = \langle strcat, strcat, fopen \rangle$,

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

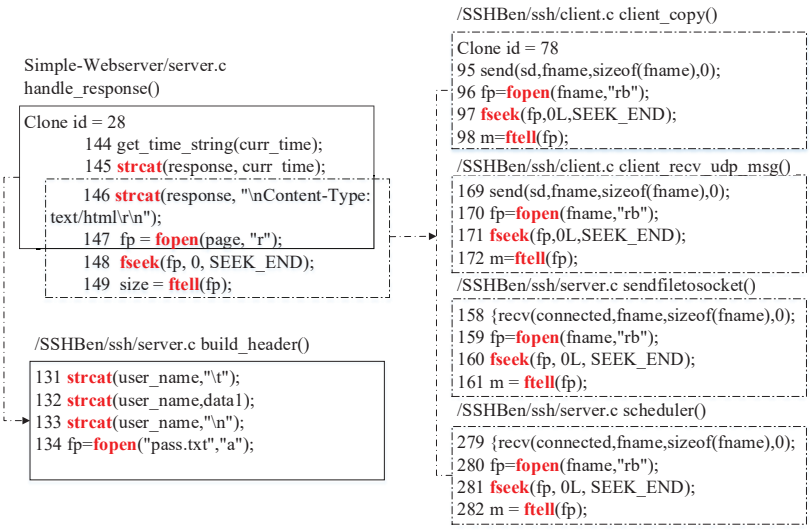


Fig. 3. An Example of Cloned Segments

- cas_{28}(SSHBen|SimpleWebserver) = $\langle strcat, strcat, strcat, fopen\rangle$,
- cas_{78}(SimpleWebserver|SSHBen) = $\langle fopen, fseek, ftell\rangle$, and
- cas_{78}(SSHBen|SimpleWebserver) = $\langle fopen, fseek, ftell, fopen, fseek, ftell, fopen, fseek, ftell, fopen, fseek, ftell\rangle$.

Since the attack surface concept is based on the number of entry and exist points (i.e., methods invoking I/O functions), we follow the similar approach to calculate the size of common attack surface by counting the number of I/O function calls across different Clone ids, with those appearing under different Clone ids counted only once. We demonstrate this through an example.

Example 3. For Clone id 78, this gives three for *Simple-Webserver* and 12 for *SSHBen*. As to Clone id 28, we have three for *Simple-Webserver* and four for *SSHBen*. Note that *fopen* is considered under both Clone ids for *Simple-Webserver*, and hence we should count it only once. Based on those discussions, we can calculate the total number of I/O function calls for both Clone ids as five for *Simple-Webserver* and 16 for *SSHBen*.

Finally, we define the *Conditional Common Attack Surface* as the ratio between the size of the common attack surface of a software application (w.r.t. to another software) and the size of its entire attack surface (i.e., the total number of I/O function calls inside that software). This ratio indicates the degree to which the software shares with others similar I/O function calls (entry/exit points).

Definition 2 (Conditional Common Attack Surface). *Given two software applications A and B with totally n clone segments, and AS_A and AS_B as the total number of I/O function calls inside A and B, respectively, the conditional common attack surface*

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

of A w.r.t B (or that of B w.r.t. A), denoted as ccas(A|B) (or ccas(B|A)), is defined as:

$$ccas(A \mid B) = \frac{|\bigcup_{i=1}^{n} cas(A \mid B)|}{AS_A}$$

$$ccas(B \mid A) = \frac{|\bigcup_{i=1}^{n} cas(B \mid A)|}{AS_B}$$

Example 4. The attack surface (i.e., the total number of I/O function calls) of *SimpleWebserver* and *SSHBen* are 16 and 182, respectively. We thus have $ccas(SSHBen \mid SimpleWebserver) = \frac{5}{16} = 0.3125$ and $ccas(SimpleWebserver \mid SSHBen) = \frac{16}{182} = 0.029$. The results show that *SSHBen* contains about 31% shared attack surface, whereas *SimpleWebserver* contains only 2.9%. By comparing a software application to many others, the developer of that application may gain useful insights from such results in terms of vulnerability discovery and security patch management.

3.2 Probabilistic Common Attack Surface Metric

The conditional common attack surface metric *ccas* is designed for evaluating one software application against others. We now take a different approach of defining a symmetric probabilistic common attack surface metric for two software applications. Such a metric can be used to estimate the amount of effort that a potential attacker may reuse while attempting to compromise both software applications. The nature of such a use case implies the metric should be symmetric.

We apply Jaccard index for this purpose, which is commonly defined as $J(A, B) = \frac{A \cap B}{A \cup B}$ and used for analyzing the similarity and diversity between the two sets. To apply this metric in our case, we need to define both the intersection and union of the attack surface of two software applications. The common attack surface defined in previous section (Definition 1) can be considered as the intersection, but such a definition is not sufficient here since it is asymmetric in nature. Instead, we will define the intersection between the attack surface of two software applications using the standard multi-set intersection operation [42], which is described below.

Definition 3 (Intersection of Multi-Sets [42]). *Given two multi-sets A = $\langle A, f\rangle$ (where f is the multiplicity function such that for any $a \in A$, $f(a)$ gives the number of occurrences of a in the multiset) and B = $\langle A, g\rangle$, then their intersection, denoted as A∩B, is the multi-set $\langle A, s\rangle$, where for all $a \in A$:*

$$s(a) = min(f(a), g(a)).$$

Example 5. Assume U ={a,a,a,b} and V = {a,a,b,b}, if we apply the multi-set operation as defined above, we have $U \cap V$ = {a,a,b}.

The union of the attack surface between two software applications can be defined as $AS_A \cup AS_B = AS_A + AS_B - cas(B \mid A) \cap cas(A \mid B)$. With both the union and intersection operations defined, we can now define the probabilistic common attack surface metric as follows.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

Definition 4 (Probabilistic Common Attack Surface Metric). *Given two software applications A and B, with their attack surface AS_A and AS_B and the common attack surface cas(B|A) and cas(A|B), respectively, the probabilistic common attack surface of A and B is defined as:*

$$pcas(A.B) = \frac{\mid cas(B \mid A) \cap cas(A \mid B) \mid}{\mid AS_A \cup AS_B \mid}$$

Example 6. The size of attack surface in ***Simple-Webserver*** and ***SSHBen*** is 16 and 182, respectively. From our previous discussions, we have $cas(SSHBen \mid Simple\text{-}Webserver) \cap cas(SimpleWebserver \mid SSHBen) = \langle strcat, strcat, fopen, fseek, ftell \rangle$ whose size is 5, and hence $pcas(SSHBen.SimpleWebserver) = \frac{5}{16+182-5} = 2.6\%$. Intuitively, this result indicates that, among all the I/O function calls, about 2.6% are shared between the two software applications. Such a result, when applied to all pairs of software applications inside a network, may allow administrators to estimate the degree of software diversity in the network from a security point of view.

4 Design and Implementation

To automate the evaluation of common attack surface between software applications, we design and implement a tool, *CASFinder*. Figure 4 depicts the architecture of *CASFinder*, which consists of three main components, the clone detection module, the source code labeling module, and the visualization module. The following describes those modules in more details.

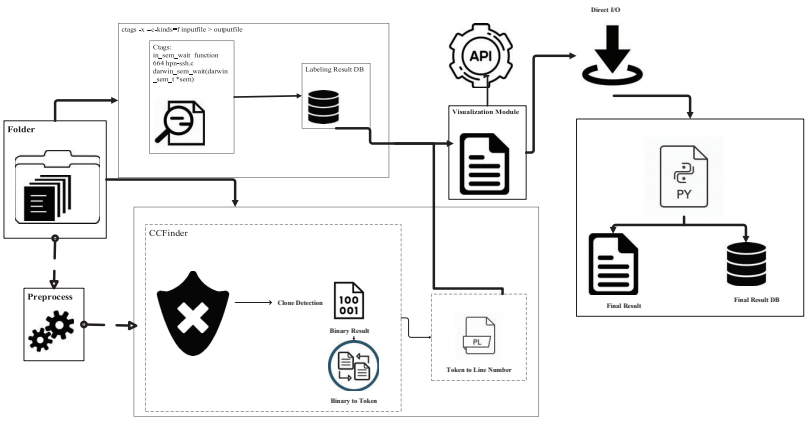


Fig. 4. The Architecture

– *The Clone Detection Module* As mentioned earlier, we choose *CCFinder* [22] as the basis of our clone detection module. The following details challenges and solutions

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

for applying *CCFinder*. First, since our tool is developed and operated under Linux, we apply only the back end of *CCFinder*. One challenge is that, since the default Linux version of *CCFinder* is designed to work on *Ubuntu 9*, the newer versions of many libraries are no longer valid for *CCFinder*. Therefore, several libraries need to be installed separately, e.g., *libboost-dev* and *libicu-dev*, which will depend on the specific version of the Linux system and can be determined based on the warnings and errors produced by *CCFinder*. Second, various parameters can be fine tuned in *CCFinder* to customize its execution mode [21]. In particular, the most important parameters include b, the minimum length of the detected code clones, and t, the minimum number of types of tokens involved. We have chosen $b = 20$ and $t = 8$ based on experiences obtained through extensive experiments. In addition, parameter w is used to determine whether *CCFinder* will perform inner-file clone detection whose results contain clones between different parts of the same software application, which is not our focus, and therefore w is set to be *f-w-g+* to focus on inter-file clones. Finally, the default output of the *CCFinder* is stored in a binary file with *.ccfd* extension. Since we do not install any front end of *CCFinder*, we apply the command *./$PATH/ccfx -p name.ccfd* to translate the *.ccfd* file into a human-readable version. The resultant file contains only the token information, which cannot be directly mapped back to the source code files. Therefore, we have developed a script, *post-prettyprint.pl* [37], to convert the token information into corresponding line numbers in the source code.

– *The Source Code Labeling Module* As mentioned above, the converted output of *CCFinder* provides only the file name and line number of the clone segments, without information needed for mapping them back to the original source code. For the purpose of generating traceable output with source code fragments, a mapping between the line number of the clone segments and the source code needs to be established. This second module is designed for this purpose by automatically retrieving a clone code segment from the source code according to the result of *CCFinder*.

– *The Visualization and CAS Calculation Module* The visualization module generates the results of clone segments. The results include clone ID, file path, function name, clone segment, start line number, and end line number. The visualized output is organized as an *XML* tree with labels. The label *contents* contains the source clone segments from *CCFinder* outputs. Label *funcname* reveals the function names corresponding to the clone segments, and label *io* contains the common I/O functions. To calculate the common attack surface, we first need to identify the I/O functions. In our experiments, we have obtained the list of I/O functions from the GNU C library [39] (glibc), which is the GNU project's implementation of C standard library, as the database for examining the entry/exit points. In total, 256 I/O functions are stored in our database, e.g., function *memcpy()* or *strcpy*, which could take user inputs as the source, and copy them directly to the memory block pointed to by the destination. Such functions have caused many serious security flaws including CVE-2014-0160 (i.e., the Heartbleed bug [7]). The final result of common attack surface is calculated based on the I/O functions shared among all software applications, and can be stored either in a file or into the database.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

5 Experiments

This section presents experimental results on applying our tool *CASFinder* to real world open source software.

5.1 Dataset

To study the common attack surface among real world software applications, we need a large amount of open-source software to apply our tool. For this purpose, we have developed a script to automatically parse the download links at the open-source software hosts. Our research shows that *GitHub* [14] provides the customized API for users to search open-source software applications with customized requirements and to download them automatically. The results are presented in json code, which contains the download link of each application together with other information. In our experiments, we have set the parameter *language* to C programs, and use parameters *q*, *sort*, and *order* to specify the query conditions and to customize the sequence of results. We have developed the script to parse the json format output from the *GitHub* automatically and to store the information of the software download link, authors, publish time, size, and other descriptions into our local database. All the download links for each software application are stored separately. Since *Github* has a limitation with respect to the maximum requests in a certain amount of time, we design the process to sleep for certain time after each query. Our experimental environment is a virtual machine running Ubuntu 14.04, with the Intel core i3-4150 CPU and 8.0GB of RAM. We have applied our tool to totally 293 different software applications belonging to seven categories. The software applications belong to several categories as follows: 32 in Databases, 62 in Web servers, 25 in ssh servers, 79 in FTP servers, 41 in TFTP servers, 6 in IMAP servers, and 48 in firewalls. Those amount to totally $\binom{293}{2} = 42778$ pairs of software applications tested using our tool in the experiments.

5.2 Cross-Category Common Attack Surface

In this section, we apply the two proposed common attack surface metrics to totally 42,778 pairs of real world software. The first set of experiments reveal the existence of common attack surface between different categories of software applications. To convert the results to a comparable scale, we have normalized the absolute value of common attack surface reported by *CASFinder* by the size of the software. Figure 5 shows the existence of common attack surface across seven categories. The percentages on top of the bars inside each figure indicate the level of common attack surface between the category mentioned in the title of the figure and all the seven categories. We can observe that common attack surface exists in all of the category combinations. For example, the *DB* category has the highest level of common attack surface inside its own category (between different software inside that category), 27.9%, and it also shares more than 9% common attack surface with any other category.

In summary, the results across all categories are shown in the heat map in Table 1 where a darker color indicates a larger CAS value between the pair of categories. A visible diagonal with the darkest color in the heat map indicates the expected trend that

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

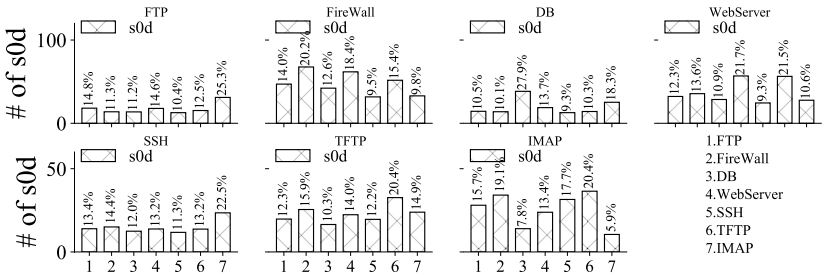


Fig. 5. Common Attack Surface across Categories

different software in the same category yield the highest level of common attack surface, most likely due to their similar functionality, except for *SSH*. In fact, the category *SSH* has the lowest level of common attack surface within its category. The reason is that the *SSH* category only contains 25 software applications, which is not sufficiently large to produce any reliable trend. Due to similar reasons, we have omitted the results from the *IMAP* category in the heat-map.

	FTP	FireWall	DB	WebServer	SSH	TFTP
FTP	18.2	13.8	13.7	17.9	12.8	15.3
FireWall	47.1	67.6	42.2	61.8	31.7	51.7
DB	14.4	13.9	38.4	18.8	12.8	14.1
WebServer	32.2	35.6	28.6	56.9	24.4	56.3
SSH	13.9	15.0	12.5	13.8	11.8	13.7
TFTP	19.8	25.5	16.5	22.4	19.6	32.6

Table 1. HeatMap for Common Attack Surface in Different Categories

After understanding the general existence of common attack surface among the seven categories of software applications, we aim to study more specific trends in our second sets of experiments. The left chart in Figure 6 shows the accumulated number of pairs of software applications in the absolute value of common attack surface. The figure depicts only the results with a nonzero value, which include totally 9,852 pairs (which amounts to about 1/8 of the total number of pairs). We can observe that the accumulated number of pairs of software applications increases quickly before the value of common attack surface reaches about 12 and afterwards the accumulation flattens out. About 20% of software share common clone segments, and 56% of the clone segments contain at least one common attack surface. The right chart in Figure 6 depicts the relationship between common attack surface and sizes of the software. We use the absolute values of common attack surface in this experiment. For the sizes, we use the normalized combined sizes $\log_{1000}(A^B)/1000$ when software A is compared with software B. We can observe that, with increasing sizes of the software, the value of common attack surface generally increases. This is as expected since the number of I/O functions would be roughly proportional to the size of the software.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

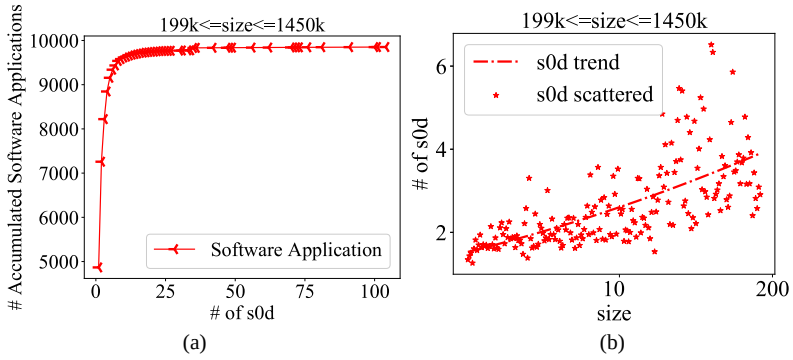


Fig. 6. CAS in Accumulated Software Application Pairs(a), CAS Trend vs Size(b)

The left chart in Figure 7 compares the average number of I/O functions and the average common attack surface over several years. The blue bars indicate the average number of I/O functions used in the software applications tested in our experiments based on the publishing year. The average number of I/O functions per software application does not have a simple trend and is used as a baseline for comparison. We can observe a clear downward trend in the average value of common attack surface over time, with software published around 2010 having a much higher value of common attack surface compared with more recent years, regardless of the number of average I/O functions. We believe this trend shows that code reusing plays a major role in common attack surface, since the trend can be easily explained by the backward nature of code reusing (i.e., programmers can only reuse older code). The right chart in Figure 7 explores the trend of the probabilistic common attack surface metric versus the size. The value of the probabilistic common attack surface metric decreases since the increase of the number of I/O functions in software applications is faster than the increase of common attack surface.

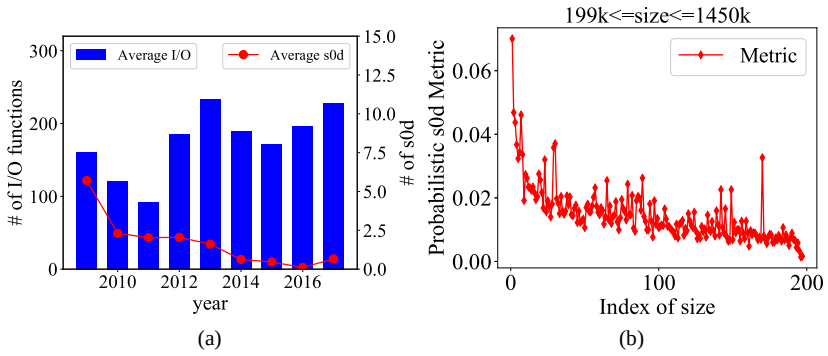


Fig. 7. CAS Trend in Years(a) and The Probabilistic CAS Metric(b)

In fact, those results match the results of existing vulnerability discovery models, which generally show that larger software applications typically have more vulnerabilities but a lower probability for having vulnerabilities per unit of software size. For

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

example, Google Chrome (with the number of lines at 14,137,145 [1]) has 1,453 vulnerabilities over nine years [8], while Apache (with the number of lines at 1,800,402) has 815 over 19 years. However, the probability of having one vulnerability per unit of software size per year is $1.15 * 10^{-3}$% for Chrome and $2.4 * 10^{-3}$% for Apache (i.e., the larger Chrome has less vulnerabilities per unit of software size).

5.3 Common Attack Surface in the Same Category

We study the trend of common attack surface between software within the same category in this section. Figure 8 depicts the common attack surface for different sizes of software in the category *WebServer* and *FTP*, respectively, represented in both scattered and trending results. The orange scattered points and the dotted line indicate the result and the red dotted line is the same trend borrowed from Figure 6for comparison. We can observe that the trend of common attack surface in both categories increase with the size, which follows a similar trend as the cross category result. However, the trend of *WebServer* increases faster than the cross-category trend, which matches the results shown in Table 1. On the other hand, the trend in the *FTP* category grows slightly slower than the cross category trend, which can be explained by the fact that *FTP* shares a large amount of common attack surface with *WebServer* and *TFTP*.

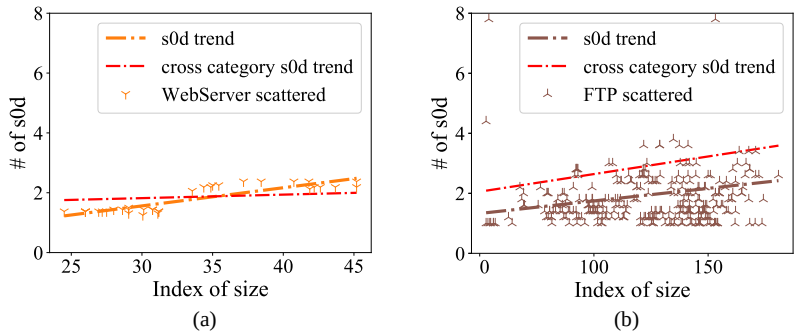


Fig. 8. Size Trend in Same Category, WebServer (a) and FTP (b)

The left chart in Figure 9 depicts the trend of common attack surface over time in the same category. Each blue bar represents the average number of I/O functions in the years in the same category of the experiments. The red line shows the average number of common attack surface in those years. Compared to Figure 7, the common attack surface in the same category has higher values, which also match the previous observations. The right chart in Figure 9 reveals the trend of the probabilistic common attack surface metric versus the size in the same category, which shows a similar trend as the cross category result, although the trend within the same category starts from a higher value around 0.20 (in contrast, the cross-category metric starts from 0.06).

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

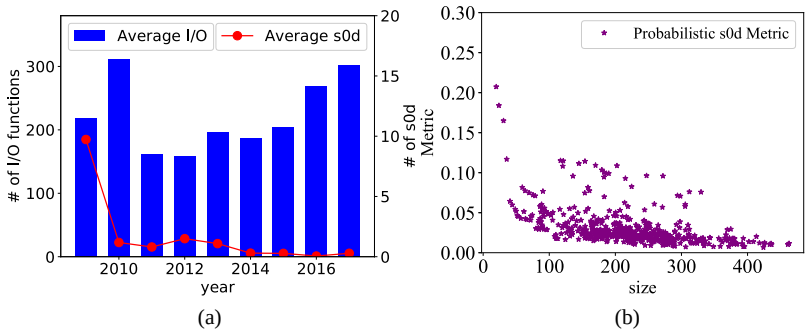


Fig. 9. Common Attack Surface Over Time and vs Size

6 Related Work

There exist extensive research on clone code detection although many of these tools are mainly for research purposes [41]. One of the popular tools in text-based clone detection is the *Dup* [2]; if two lines of code are identical after removing all whitespaces and comments, they are assigned as clone codes; the longest line matches are the output, but the minimum length of the reported code can be customized according to different needs. Another well-known approach [20] is applying the fingerprint in order to identify the redundancy on a substring of the source code. The fingerprinting calculation uses KARP-Rabins string matching approach [24, 25] to calculate the length of all n substrings. Ducasse developed [9] *duploc* which was designed to be a parsing free, language-independent tool which first reads the source file and sequences of the lines, then removes all comments and whitespace to create a set of condensed lines; afterward, a comparison is made based on the hash result, where scatter-plots indicate the visualization of a cloned result. Token-based clone detection is also one of the widely applied methods. One of the representative tools in token-based detection is *CCFinder* [22], which is applied in our work. Bakers Dup [2, 3] implements a similar approach as *CCFinder*. The detection process begins by tokenizing the source code, then using a suffix-tree algorithm to compare tokens. Unlike *CCFinder*, *Dup* does not apply transformation, but rather consistently renames the identifier. Raimar Falke [29] develops a tool called *iclones* [15], which uses suffix-trees to find clones in abstract syntax trees, which can operate in linear time and space. CP-Miner [31] as a well-designed token-based clone detector, uses frequent subsequence mining algorithms to detect tokenized segments. RTF [5] is a token-based clone detector that uses string algorithms for efficient detection; rather than using the more common suffix-tree, it utilizes more memory-efficient suffix array.

One of the leading tools using AST-based algorithm is the *CloneDR* developed by Baxter [6] which can detect exact and near-miss clone through applying hashing and dynamic algorithm. The *ccdiml* [38] developed by Bauhaus is similar to the *CloneDR* in the way of dealing with hash and code sequences, but instead of using AST, it applies IML algorithm in the comparing process. David and Nicholas [13] develop a tool named *Sim* which uses a standard lexical analyzer to generate a parsing-tree of two

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

given software applications. The code similarity is determined by applying the maximum common subsequence and dynamic programming. One of the leading PDG-based tools is PDG-DUP presented by Komondoor and Horwit [26] and Komondoor and Horwitz's PDG-DUP [26] is another leading PDG-based detection tool, which identifies clones together and keeping the semantics of the source code to reflect software. As to metric-based clone detection, in [36] Mayrand uses the tool *Darix* to generate the metric and the clone identification is based on four values, which are name, layout, expression and control flow [36]. Kontogiannis [27] uses Markov models to compute the dissimilarity of the code by applying the abstract pattern matching. Five widely used metrics are applied in a direct comparison in [28]. There are also some other approaches that using hybrid clone detections. In [29], the authors apply the suffix trees to find clones in AST; this approach can find clones in linear time and space.

The concept of attack surface is originally proposed for specific software, e.g., Windows, and requires domain-specific expertise to formulate and implement [16]. Later on, the concept is generalized using formal models and becomes applicable to all software [34]. Furthermore, it is refined and applied to large scale software, and its calculation can be assisted by automatically generated call graphs [33, 32]. Attack surface has attracted significant attentions over the years. It is used as a metric to evaluate Android's message-passing system [23], in kernel tailing [30], and also serves as a foundation in Moving Target Defense, which basically aims to change the attack surface over time so to make attackers' job harder [18, 17]. The study on automating the calculation of attack surface is another interesting domain, e.g., COPES uses static analysis from bytecode to calculate attack surface and to secure permission-based software [4]. Stack traces from user crash reports is used to approximate attack surface automatically [43]. The correlation between attack surface and vulnerabilities has also been investigated, such as using attack surface entry points and reachability to assess the risk of vulnerability [46]. A study about the relationship between attack surface and the vulnerability density is given in [45], although the result is only based on two releases of Apache HTTP Server. Despite such interest in attack surface, to the best of our knowledge, the common attack surface between different software has attracted little attention.

7 Conclusion

In this paper, we have defined the concept of common attack surface and implemented an automated tool for evaluating the common attack surface between given software applications. We have conducted experiments on real open source software and examined the common attack surface both within and between software categories. Our results have shown common attack surface to be pervasive among software. Our work still has some limitations which will lead to our future work. First, since we rely on *CCFinder* our tool also inherits its limitations, and one future direction is to explore other clone detection tools. Second, we have focused on entry/exit points of attack surface, and one future direction is to also consider channels and untrusted data items. Third, we have focused on the C language in this work, and extending it to other languages with different entry and exit libraries is an interesting future direction. Finally, we plan to extend the effort on correlating between common attack surface and known vulnerabilities.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

Disclaimer

Commercial products are identified in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the identified products are necessarily the best available for the purpose.

References

1. Open hub. `https://www.openhub.net/`, 2017.
2. Brenda S Baker. A program for identifying duplicated code. *Computing Science and Statistics*, pages 49–49, 1993.
3. Brenda S Baker. On finding duplication and near-duplication in large software systems. In *Reverse Engineering, 1995., Proceedings of 2nd Working Conference on*, pages 86–95. IEEE, 1995.
4. Alexandre Bartel, Jacques Klein, Yves Le Traon, and Martin Monperrus. Automatically securing permission-based software by reducing the attack surface: An application to android. In *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering*, pages 274–277. ACM, 2012.
5. Hamid Abdul Basit and Stan Jarzabek. Efficient token based clone detection with flexible tokenization. In *Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering*, pages 513–516. ACM, 2007.
6. Ira D Baxter, Andrew Yahin, Leonardo Moura, Marcelo Sant'Anna, and Lorraine Bier. Clone detection using abstract syntax trees. In *Software Maintenance, 1998. Proceedings., International Conference on*, pages 368–377. IEEE, 1998.
7. Marco Carvalho, Jared DeMott, Richard Ford, and David A Wheeler. Heartbleed 101. *IEEE security & privacy*, 12(4):63–67, 2014.
8. CVE Community. Common vulnerabilities and exposures. `https://cve.mitre.org/`, 1999.
9. Stéphane Ducasse, Matthias Rieger, and Serge Demeyer. A language independent approach for detecting duplicated code. In *Software Maintenance, 1999.(ICSM'99) Proceedings. IEEE International Conference on*, pages 109–118. IEEE, 1999.
10. Zakir Durumeric, James Kasten, David Adrian, J Alex Halderman, Michael Bailey, Frank Li, Nicolas Weaver, Johanna Amann, Jethro Beekman, Mathias Payer, et al. The matter of heartbleed. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 475–488. ACM, 2014.
11. Seyed Mohammad Ghaffarian and Hamid Reza Shahriari. Software vulnerability analysis and discovery using machine-learning and data-mining techniques: A survey. *ACM Computing Surveys (CSUR)*, 50(4):56, 2017.
12. AK Ghosh, D Pendarakis, and WH Sanders. Moving target defense co-chairs report-national cyber leap year summit 2009. *Tech. Rep., Federal Networking and Information Technology Research and Development (NITRD) Program*, 2009.
13. David Gitchell and Nicholas Tran. Sim: a utility for detecting similarity in computer programs. In *ACM SIGCSE Bulletin*, volume 31, pages 266–270. ACM, 1999.
14. GitHub.Inc. A web-based hosting service for version control using git. `https://github.com`.
15. Nils Göde and Rainer Koschke. Incremental clone detection. In *Software Maintenance and Reengineering, 2009. CSMR'09. 13th European Conference on*, pages 219–228. IEEE, 2009.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

16. M. Howard, J. Pincus, and J. Wing. Measuring relative attack surfaces. In *Workshop on Advanced Developments in Software and Systems Security*, 2003.
17. S. Jajodia, A.K. Ghosh, V. S. Subrahmanian, V. Swarup, C. Wang, and X.S. Wang. *Moving Target Defense II: Application of Game Theory and Adversarial Modeling*. Springer, 2012.
18. S. Jajodia, A.K. Ghosh, V. Swarup, C. Wang, and X.S. Wang. *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*. Springer, 1st edition, 2011.
19. Sushil Jajodia, Anup K Ghosh, Vipin Swarup, Cliff Wang, and X Sean Wang. *Moving target defense: creating asymmetric uncertainty for cyber threats*, volume 54. Springer Science & Business Media, 2011.
20. J Howard Johnson. Substring matching for clone detection and change tracking. In *ICSM*, volume 94, pages 120–126, 1994.
21. Toshihiro Kamiya. Tutorial of cli tool ccfx. `http://www.ccfinder.net/doc/10.2/en/tutorial-ccfx.html`, 2008.
22. Toshihiro Kamiya, Shinji Kusumoto, and Katsuro Inoue. Ccfinder: a multilinguistic token-based code clone detection system for large scale source code. *IEEE Transactions on Software Engineering*, 28(7):654–670, 2002.
23. David Kantola, Erika Chin, Warren He, and David Wagner. Reducing attack surfaces for intra-application communication in android. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pages 69–80. ACM, 2012.
24. Richard M Karp. Combinatorics, complexity, and randomness. *Commun. ACM*, 29(2):97–109, 1986.
25. Richard M Karp and Michael O Rabin. Efficient randomized pattern-matching algorithms. *IBM Journal of Research and Development*, 31(2):249–260, 1987.
26. Raghavan Komondoor and Susan Horwitz. Using slicing to identify duplication in source code. In *International Static Analysis Symposium*, pages 40–56. Springer, 2001.
27. K Kontogiannis, M Galler, and R DeMori. Detecting code similarity using patterns. In *Working Notes of 3rd Workshop on AI and Software Engineering*, volume 6, 1995.
28. Kostas A Kontogiannis, Renator DeMori, Ettore Merlo, Michael Galler, and Morris Bernstein. Pattern matching for clone and concept detection. *Automated Software Engineering*, 3(1-2):77–108, 1996.
29. Rainer Koschke, Raimar Falke, and Pierre Frenzel. Clone detection using abstract syntax suffix trees. In *Reverse Engineering, 2006. WCRE'06. 13th Working Conference on*, pages 253–262. IEEE, 2006.
30. Anil Kurmus, Reinhard Tartler, Daniela Dorneanu, Bernhard Heinloth, Valentin Rothberg, Andreas Ruprecht, Wolfgang Schröder-Preikschat, Daniel Lohmann, and Rüdiger Kapitza. Attack surface metrics and automated compile-time os kernel tailoring. In *NDSS*, 2013.
31. Zhenmin Li, Shan Lu, Suvda Myagmar, and Yuanyuan Zhou. Cp-miner: Finding copy-paste and related bugs in large-scale software code. *IEEE Transactions on software Engineering*, 32(3):176–192, 2006.
32. P. Manadhata and J. Wing. An attack surface metric. Technical Report CMU-CS-05-155, 2005.
33. P. Manadhata and J. Wing. An attack surface metric. *IEEE Trans. Softw. Eng.*, 37(3):371–386, May 2011.
34. Pratyusa Manadhata and Jeannette Wing. Measuring a system's attack surface. Technical Report CMU-CS-04-102, 2004.
35. Pratyusa K Manadhata and Jeannette M Wing. An attack surface metric. *IEEE Transactions on Software Engineering*, 37(3):371–386, 2011.
36. Jean Mayrand, Claude Leblanc, and Ettore Merlo. Experiment on the automatic detection of function clones in a software system using metrics. In *icsm*, volume 96, page 244, 1996.
37. Petersenna. Ccfinder core. `https://github.com/petersenna/ccfinderx-core`.

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

38. Aoun Raza, Gunther Vogel, and Erhard Plödereder. Bauhaus-a tool suite for program analysis and reverse engineering. In *Ada-Europe*, volume 4006, pages 71–82. Springer, 2006.
39. Trevis Rothwell. The gnu c reference manual. `https://www.gnu.org/software/gnu-c-manual/`, 2006.
40. Chanchal K Roy, James R Cordy, and Rainer Koschke. Comparison and evaluation of code clone detection techniques and tools: A qualitative approach. *Science of computer programming*, 74(7):470–495, 2009.
41. Chanchal Kumar Roy and James R Cordy. A survey on software clone detection research. *Queens School of Computing TR*, 541(115):64–68, 2007.
42. Apostolos Syropoulos. Mathematics of multisets. In *Workshop on Membrane Computing*, pages 347–358. Springer, 2000.
43. Christopher Theisen, Kim Herzig, Patrick Morrison, Brendan Murphy, and Laurie Williams. Approximating attack surfaces with stack traces. In *Proceedings of the 37th International Conference on Software Engineering-Volume 2*, pages 199–208. IEEE Press, 2015.
44. David A. Wheeler. A program that examines c/c++ source code and reports possible security weaknesses. `https://www.dwheeler.com/flawfinder/`.
45. Awad A Younis and Yashwant K Malaiya. Relationship between attack surface and vulnerability density: A case study on apache http server. In *Proceedings on the International Conference on Internet Computing (ICOMP)*, page 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.
46. Awad A Younis, Yashwant K Malaiya, and Indrajit Ray. Using attack surface entry points and reachability analysis to assess the risk of software vulnerability exploitability. In *High-Assurance Systems Engineering (HASE), 2014 IEEE 15th International Symposium on*, pages 1–8. IEEE, 2014.
47. Mengyuan Zhang, Lingyu Wang, Sushil Jajodia, Anoop Singhal, and Massimiliano Albanese. Network diversity: a security metric for evaluating the resilience of networks against zero-day attacks. *IEEE Transactions on Information Forensics and Security*, 11(5):1071–1086, 2016.

Appendix

Common Attack Surface and Vulnerabilities

We study the correlation between the common attack surface of two software applications and their shared vulnerabilities. Although, as mentioned earlier in Section 2.2, the concept of attack surface is not intended as a one-to-one mapping to actual vulnerability, the size of attack surface can still provide a rough indicator for the relative abundance of vulnerabilities, since entry and exit points represent the interfaces exposed by the software for accepting inputs from (or sending outputs to) the outside environment. Consequently, the common attack surface may also indicate shared vulnerabilities. Therefore, we study this correlation through experiments.

To evaluate the correlation between common attack surface and vulnerabilities, we examine pairs of software applications with respect to the results of a vulnerability scanner called *flawfinder* [44]. *Flawfinder* is an open-source tool that can be used to scan C and C++ source code and report potential vulnerabilities [44]. It is regarded as an effective tool for detecting misused functions with ranked risks. For the purpose

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

of verifying the relationship between common attack surface and vulnerabilities, we manually compare our results with the results of *flawfinder*.

Our findings indicate that common vulnerabilities may indeed to some extent be correlated with common attack surface. For example, we examined the two software *SSH* and *simple-webserverche*, which are of the category *SSH* and *Webserver* applications, respectively. In *SSH*, the main file uses function *strcat* to copy data to an internal parameter *user_name*, and those data are applied without any boundary check. The same thing happens in the application *simple-webserverche* where a file named *server.c* calls function *handle_response()* to apply function *strcat*. The source parameter *curr_time* is applied before the boundary checking. Our tool successfully detects these code fragments as a common attack surface, while *flawfinder* reports that both have the potential to lead to similar *buffer overflow* vulnerabilities.

To further evaluate the extent of such correlation, we compare the outputs of *flawfinder* and the results of our tool, and Table 2 shows the level of correlation between the two. As the results show, in every category of software applications, there exist a certain percentage of vulnerabilities which correlate to the common attack surface.

FTP	4.07%	FireWall	3.74%	DB	3.40%	WebServer	4.08%	SSH	3.37%	TFTP	3.10%	IMAP	6.52%

Table 2. Percentage of Detected Vulnerabilities Which Correlate to Reported Common Attack Surface

Zhang, Meng; Xin, Yue; Wang, Lingyu; Jajodia, Sushil; Singhal, Anoop.
"CASFinder: Detecting Common Attack Surface."
Paper presented at 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, United States. July 15, 2019 - July 18, 2019.

Temperature dependence of nonlinearity in high-speed, high-power photodetectors

Josue Davila-Rodriguez[1,*], Holly Leopardi[1], Tara M. Fortier[1], Xiaojun Xie[2], Joe C. Campbell[2], James Booth[3], Nathan Orloff[3],Scott A. Diddams[1], and Franklyn Quinlan[1,†]
[1]Time and Frequency Division, NIST
Boulder, CO, USA
[2]Department of Electrical and Computer Engineering, University of Virginia
Charlottesville, VA, USA
[3]Communications Technology Laboratory, NIST
Boulder, CO, USA
[*]josuedavila@gmail.com, [†]franklyn.quinlan@nist.gov

***Abstract*—We present an experimental study of the nonlinearity of modified uni-traveling carrier (MUTC) photodiodes at cryogenic temperatures. At 120 K, the amplitude-to-phase (AM-to-PM) conversion nonlinearity is reduced by up to 10 dB, resulting in nearly 40 dB AM-to-PM rejection over a broad photocurrent range.**

Keywords— High speed photodiodes; Low noise photonic microwave generation

I. Introduction

Fast photodiodes enable the conversion of optically carried microwaves into the electrical domain, and have therefore become key to multiple applications such as microwave photonics [1], synchronization of large-scale science facilities [2] and generation of low-noise stable microwaves from optical references via optical frequency division (OFD) [3-6]. Among these applications, perhaps the most demanding is optical frequency division, since it requires high linearity, efficiency and power handling to provide sufficiently low residual noise [6] and long term stability [7] to support high fidelity division of optical signals to the microwave domain. Photodiodes based on MUTC structures have demonstrated state-of-the-art performance in each of these metrics [8].

Continued improvement of photodetector performance requires deeper understanding of their physics and limitations. Insight into device operation may be obtained by operating at low temperatures, especially as it relates to the bandgap energy and electron-phonon interaction. Furthermore, cryogenic microwave systems such as Josephson junction circuits could benefit from microwave regeneration directly in the cryogenic environment [9]. In this case, a single optical fiber could replace several coaxial cables, reducing cost, complexity and heat load. Previous studies have shown that photodiodes retain their high-speed operation at low temperatures [10] and that the dark current is significantly reduced [11]. However, we are not aware of any investigations of the nonlinearity or noise of high speed photodetectors in cryogenic environments. Here we present measurements of AM-to-PM conversion in MUTC devices at temperatures varying from 300 K down to 120 K. Our preliminary results show an improvement of up to 10 dB in the linearity of the MUTC photodiodes.

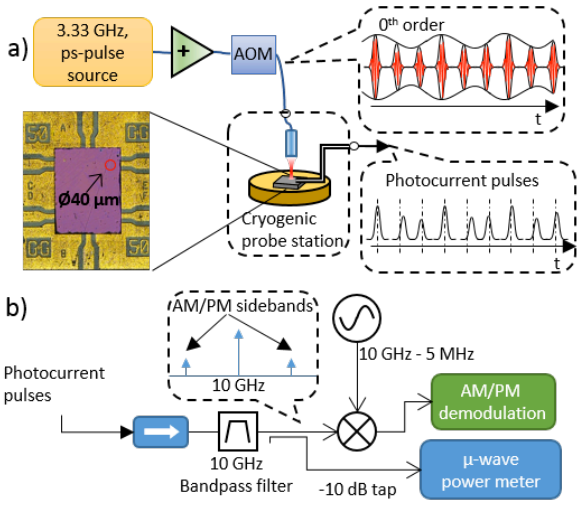


Fig. 1. Measurement setup. **a)** A pulse-train from an Er:fiber mode-locked laser is interleaved up to a repetition frequency of 3.33 GHz and a final interleaving stage generates pulse pairs separated by 100 ps. ~150 kHz amplitude modulation is written unto the beam using the 0th order of an acousto-optic modulator (AOM). **b)** The AM-to-PM demodulation is achieved by first mixing the 10 GHz harmonic down to 5 MHz and then using a FFT analyzer to demodulate the AM and PM quadratures of the 150 kHz tone. **Inset**: Photograph of the MUTC diodes, flip-chip bonded on an aluminum nitride substrate with patterned co-planar waveguides for microwave transmission. The active area of the device under test is shown in the red circle.

II. Experiments and Results

The diodes used for this study are 40 µm diameter MUTC devices as demonstrated in [12]. The photodiode is illuminated with ~1 ps pulses at 1550 nm obtained from an erbium:fiber mode-locked laser. The laser has 208.33 MHz repetition rate and is followed by a 4-stage pulse interleaver that produces a 3.33 GHz pulse-train. A final interleaving stage produces pulse pairs separated by 100 ps to maximize the power contained in the 10 GHz harmonic of the microwave spectrum. This pulse train is then delivered through a 70-m fiber link to the laboratory where the cryogenic probe station is located. At the remote site, dispersion compensation is performed with normal dispersion fiber and the pulses are subsequently amplified with

Davila-Rodriguez, Josue; Leopardi, Holly; Fortier, Tara; Xie, Xiaojun; Campbell, Joe; Booth, James; Orloff, Nathan; Diddams, Scott; Quinlan, Franklyn.
"Temperature dependence of nonlinearity in high-speed, high-power photodetectors."
Paper presented at 2017 IEEE Photonics Conference, Orlando, FL, United States. October 1, 2017 - October 5, 2017.

an erbium-doped fiber amplifier. Autocorrelation traces confirmed the picosecond pulse width delivered through a fiber optic vacuum feedthrough.

To study the photodiodes at cryogenic temperatures, we place them in a cryogenic probe station fitted with both fiber optic and microwave feedthroughs. The pulses are focused on the photodiode with a fiber-pigtailed graded index lens inside the vacuum chamber. Lateral alignment is optimized by maximizing the dc photocurrent. The focusing condition has been systematically explored, as the diode's nonlinearity is aggravated by tight focusing [13]. We observe that the diode's nonlinearity increases for beam waists much smaller than the size of the active area. For best AM-to-PM performance, we ensure that the beam size is slightly larger than the active area to overfill it. Such configuration has the advantage of maximizing the diode's linearity with the trade-off of a slight reduction in the overall responsivity.

To analyze the AM-to-PM conversion we use a standard technique [14, 15], shown in Fig 1 (b). A small amount of 150 kHz amplitude modulation is written on the optical pulse-train by modulating the driving power of an acousto-optic modulator. AM-to-PM conversion in the photodetector converts a fraction of this modulation to the phase of the generated 10 GHz microwave. To measure the resulting phase modulation, the 10 GHz signal is first mixed down to 5 MHz with a microwave synthesizer and then demodulated with an FFT analyzer. The AM-to-PM conversion is given by ratio of the phase modulation to the imparted amplitude modulation (radians per fractional optical power change).

We have measured both the microwave power in the 10 GHz harmonic and the AM-to-PM conversion in the photodiode as a function of photocurrent for several temperatures between 300 K and 120 K and reverse bias voltages between 4 and 8 V. The results for 7 V bias are shown in Fig. 2. We observe that while the null in the AM-to-PM conversion coefficient does not shift significantly with temperature, the nonlinearity over a broad photocurrent range is reduced at lower temperatures. Reasons for the improved linearity are currently under investigation.

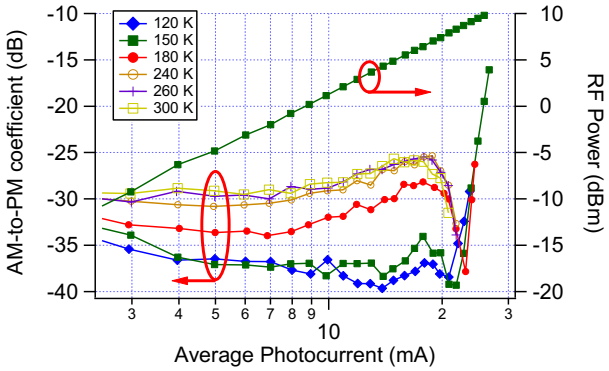


Fig. 2. **(left)** AM-to-PM conversion measured in the photodiode at 7 V bias. The null in the AM-to-PM conversion does not shift significantly, but the non-linearity at low photocurrent is significantly reduced at lower temperatures, extending the useability range of the photodiode. **(right)** 10 GHz microwave power at the bias tee at 150 K. Power levels do not change significantly with temperature.

III. Outlook

We have begun studying MUTC photodiodes at cryogenic temperatures and have found that the linearity can improve. Additionally, we intend to study the noise performance of these photodiodes to elucidate the contribution of the various mechanisms that contribute to the phase noise added by the photodiode in optically-derived low noise microwave signals.

Acknowledgements. The authors wish to thank M. Hutchinson and V. J. Urick for helpful discussions.

Funding. This work was funded by the Naval Research Laboratory, the DARPA PULSE program and NIST. This work is a contribution of an agency of the US government and not subject to copyright in the USA.

References

[1] V. J. Urick, F. Bucholtz, J. D. McKinney, P. S. Devgan, A. L. Campillo, J. L. Dexter, and K. J. Williams, "Long-Haul Analog Photonics," J. Lightwave Technol. **29**, 1182-1205 (2011)

[2] J. Kim, J. A. Cox, J. Chen, and F. X. Kaertner, "Drift-free femtosecond timing synchronization of remote optical and microwave sources," Nat. Photon. **2**, pp. 733-736 (2008)

[3] T. M. Fortier, M. S. Kirchner, F. Quinlan, J. Taylor, J. C. Bergquist, T. Rosenband, N. Lemke, A. Ludlow, Y. Jiang, C. W. Oates, and S. A. Diddams, "Generation of ultrastable microwaves via optical frequency division," Nat. Photon. **5**, 425–429 (2011).

[4] T. M. Fortier, F. Quinlan, A. Hati, C. Nelson, J. A. Taylor, Y. Fu, J. Campbell, and S. A. Diddams, "Photonic microwave generation with high- power photodiodes," Opt. Lett. **38**, 1712–1714 (2013).

[5] X. Xie, R. Bouchand, D. Nicolodi, M. Giunta, W. Hänsel, M. Lezius, A. Joshi, S. Datta, C. Alexandre, M. Lours, P.-A. Tremblin, G. Santarelli, R. Holzwarth, and Y. Le Coq, "Photonic microwave signals with zeptosecond level absolute timing noise," Nat. Photon. **11**, 44-47 (2016).

[6] F. Quinlan, T. M. Fortier, H. Jiang, A. Hati, C. Nelson, Y. Fu, J. C. Campbell, and S. A. Diddams, "Exploiting shot noise correlations in the photodetection of ultrashort optical pulse trains," Nat. Photon. **7**, 290–293 (2013).

[7] F.N. Baynes, F. Quinlan, T.M. Fortier, Q. Zhou, A. Beling, J.C. Campbell, and S.A. Diddams, "Attosecond timing in optical-to-electrical conversion," Optica **2**, 141-146 (2015)

[8] X. Xie, Q. Zhou, K. Li, Y. Shen, Q. Li, Z. Yang, A. Beling, and J. C. Campbell, "Improved power conversion efficiency in high-performance photodiodes by flip-chip bonding on diamond," Optica **1**, 429-435 (2014)

[9] B. Van Zeghbroeck, "Optical data communications between Josephson-junction circuits and room-temperature electronics," IEEE Trans. Appl. Sup., **3**, 2881 (1993)

[10] H. Ito, T. Furuta, S. Kodama, K. Yoshino, T. Nagatsuma and Z. Wang, "10 Gbit/s operation of uni-travelling-carrier photodiode module at 2.6 K," Electronics Lett., **44**, pp. 149-150 (2008).

[11] Y. M. Zhang, V. Borzenets, N. Dubash, T. Reynolds, Y. G. Wey and J. Bowers, "Cryogenic performance of a high-speed GaInAs/InP p-i-n photodiode," J. of Lightw. Tech., **15**, pp. 529-533 (1997).

[12] X. Xie, J. Zang, A. Beling, and J. Campbell, "Characterization of Amplitude Noise to Phase Noise Conversion in Charge-compensated Modified Uni-Travelling Carrier Photodiodes," in J. of Lightw. Tech. doi: 10.1109/JLT.2016.2641967

[13] A. Joshi, S. Datta and D. Becker, "GRIN Lens Coupled Top-Illuminated Highly Linear InGaAs Photodiodes," Photon. Tech. Lett., **20**, pp. 1500-1502 (2008)

[14] J. Taylor, S. Datta, A. Hati, C. Nelson, F. Quilan, A. Joshi, and S. A. Diddams, "Characterization of Power-to-Phase Conversion in High-Speed P-I-N Photodiodes," IEEE Photon. Journal, **3**, pp. 140-151 (2011)

[15] W. Zhang, T. Li, M. Lours, S. Seidelin, G. Santarelli, and Y. Le Coq, "Amplitude to phase conversion of InGaAs pin photo-diodes for femtosecond lasers microwave signal generation," Appl. Phys. B **106**: 301 (2012)

Davila-Rodriguez, Josue; Leopardi, Holly; Fortier, Tara; Xie, Xiaojun; Campbell, Joe; Booth, James; Orloff, Nathan; Diddams, Scott; Quinlan, Franklyn.
"Temperature dependence of nonlinearity in high-speed, high-power photodetectors."
Paper presented at 2017 IEEE Photonics Conference, Orlando, FL, United States. October 1, 2017 - October 5, 2017.

REFERENCE DATA SET AND VARIABILITY STUDY FOR HUMAN SKIN REFLECTANCE

Cooksey, C.C., Allen, D.W., Tsai, B.K.
National Institute of Standards and Technology, Gaithersburg, MD, USA
catherine.cooksey@nist.gov

DOI 10.25039/x46.2019.PO065

Abstract

The optical properties of human skin have been of interest to researchers for some time. Their interest is based on a need to know for a variety of different applications, which range from spectral imaging for automated or stand-off detection, non-invasive clinical diagnostic tools, improved models for understanding light propagation, to colour-based applications, such as reproduction of skin colour in photography and printing and colour-matching in cosmetics industries. Here we present a summary of a study that aimed to create a large data set of high-quality, human-skin reflectance spectra and quantify the variability of the resulting data set.

Keywords: Reflectance, skin colour, skin colour variation, spectral

1 Motivation

In 2012, NIST began collecting the reflectance spectra of human skin over a broad spectral range. The rapidly expanding application of spectral imaging to numerous measurement challenges involving human skin highlighted the importance of establishing a traceable, reference data set of human skin. Published literature of the optical properties of human skin for the visible region is prevalent, but data is sparse in the ultraviolet and shortwave infrared. Furthermore, spectra published up to that time typically involved only a small number of participants. This proceedings paper describes a large reference data set of human skin reflectance spectra covering the ultraviolet, visible, near-infrared, and shortwave infrared spectral regions. Several measures of variability are applied to the data, including a colour analysis.

2 Reference Data Set

NIST published a reference data set of 100 reflectance spectra of human skin, spanning the wavelength range from 250 nm to 2500 nm (Cooksey et al., 2017). The spectra are directly traceable to the national scale for directional-hemispherical reflectance factor. The methods of data collection, processing, and estimating uncertainties is described extensively in Cooksey et al., 2017. It is briefly summarized below.

2.1 Human Subjects

Volunteers participated in the human subject study, Reflectance Measurements of Human Skin, which was approved by the NIST Institutional Review Board reviewed and approved. All subjects provided written informed consent. Subjects were not selected based on age, gender, or ethnicity, nor was this information collected as metadata. Subjects were not excluded for the use of sunscreen, body lotion, or medication, or for the presence of freckles, moles, tattoos, or skin conditions or disorders.

Each subject participated in one measurement session, which consisted of acquiring reflectance measurements of the test area. The test area was a 25 mm diameter circle located on the inside of the subject's right forearm.

2.2 Reflectance Measurement

The reflectance measurements were acquired using a commercially available spectrophotometer equipped with a 150 mm integrating sphere. The subject positioned his/her

Cooksey, Catherine; Allen, David; Tsai, Benjamin.
"Reference data set and variability study for human skin reflectance."
Paper presented at 29th Quadrennial Session of the International Commission on Illumination (CIE), Washington, D.C., United States. June 14, 2019 - June 22, 2019.

bare, right forearm next to the sample port of the integrating sphere during measurement acquisition. For each subject, 3 scans were collected. The subject was permitted to rest his/her arm, removing it from the testing position, in between each scan. The parameters for the measurements are provided in Table 1.

Table 1 – Measurement parameters

Wavelength Range (nm)	Wavelength Interval (nm)	Spectral bandwidth (nm)	Source	Detector
250 to 319	3	3	Deuterium lamp	Photomultiplier tube
319 to 860	3	3	Tungsten-halogen lamp	Photomultiplier tube
860 to 2500	3	≤ 20 (variable)	Tungsten-halogen lamp	Indium-gallium-arsenide

2.3 Calculation of Reflectance Factors and Uncertainties

The resulting data was processed to produce spectra of reflectance factor values for a measurement geometry with an 8° angle of incidence and hemispherical detection, abbreviated as 8°/di. The reflectance factor values from the 3 scans were averaged to obtain the final 8°/di spectral reflectance factors for each participant.

The reference standard used for these measurements was sintered polytetrafluoroethylene (PTFE). The spectral reflectance factors for this standard are traceable to the scale for spectral reflectance factor of pressed PTFE, which was established using the absolute method of Van den Akker (Venable, 1977) in the NIST Spectral Tri-function Automated Reference Reflectometer (STARR) facility (NIST, 1998).

The estimated measurement uncertainties for the reflectance measurements are calculated according to the procedures outlined in NIST, 1994. Sources of uncertainty are the 8°/di spectral reflectance factor of the sintered PTFE standard, the sphere geometry, the wavelength, and random effects or repeatability. The evaluated contributions of each source and the expanded uncertainty (k = 2) are given in Table 2. These uncertainties are representative of the manner in which the instrument was used in this study.

Table 2 – Instrument uncertainty and its components

Source of Uncertainty	Standard Uncertainty	Uncertainty Contribution
Reflectance Standard	0.002	0.002
Geometry	0.001	0.001
Wavelength	0.3 nm	0.0008
Repeatability	0.0003	0.0003
		Expanded Uncertainty (k=2)
		0.0048

3 Data Analysis

The variability of the reflectance spectra for the 100 human subjects that participated in the study can be described in several ways (Cooksey, 2013; Cooksey, 2014; Cooksey, 2015). Figure 1 shows selected spectra from the set. The spectrum depicted by the black dashed curve is the representative of the mean spectral reflectance values of all subjects participating in the study. The representative spectrum was selected from among the set of reflectance spectra based on its similarity to the mean spectrum using the following equation:

Cooksey, Catherine; Allen, David; Tsai, Benjamin.
"Reference data set and variability study for human skin reflectance."
Paper presented at 29th Quadrennial Session of the International Commission on Illumination (CIE), Washington, D.C., United States. June 14, 2019 - June 22, 2019.

$$\theta_i = \cos^{-1}\left(\frac{{S_m}^T S_i}{\|S_m\|\|S_i\|}\right) \quad (1)$$

where

θ_i is the difference between spectra (reported in radians);
S_m is the mean spectrum of reflectance factors for the full set of scans acquired (300 scans);
S_i is an individual spectrum from the full set of scans.

The selected spectrum has the smallest resulting angle and is considered the closed match to the mean spectrum. Selecting a representative spectrum form the overall set prevented the loss of spectral features that would have resulted from averaging the small shifts inherent in the spectral variability.

The variability observed for the full set of scans was calculated using the standard deviation and is referred to as the population variability. It is depicted in Figure 1 by the grey shaded area about the representative of the mean. The spectra (grey solid) representing the total range of observed reflectance factors are also shown in Figure 1.

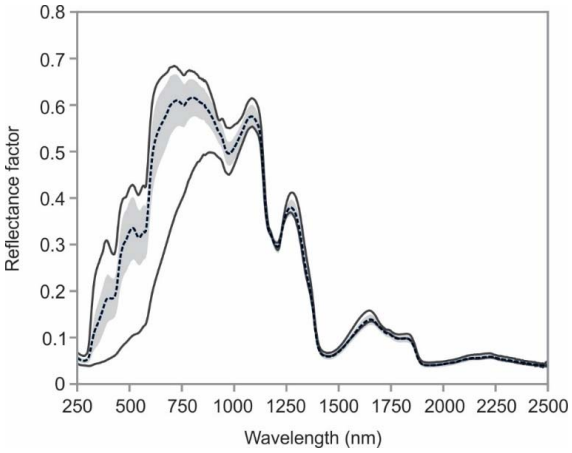


Figure 1 – The reflectance spectrum of the representative of the mean (black dashed) with grey shaded area representing the population variability for all subjects. Representative spectra (grey solid) show the range of variation in reflectance factors observed in the data set.

The greatest variation in spectral features is observed in the ultraviolet and visible spectral regions. In this spectral region, the incident light probes the epidermis, where the spectral response due to various blood-borne pigments is tempered by the absorption of light due to the presence of melanin. In contrast, there is significantly less variability observed in the near- and shortwave infrared where the incident light probes the hydrated dermal layer. The “valleys” observed in this region correspond to water absorption peaks.

Figure 2 plots the population variability with the instrument uncertainty (see Table 2) and the subject variability for selected subjects. The subject variability is the standard deviation of three scans of each subject, and it represents the dynamic nature of human skin observed for each subject. Overall, the population variability is the most significant source of uncertainty for skin’s spectral response from the ultraviolet to the near-infrared regions. In the shortwave infrared, the instrument uncertainty is a dominant source of uncertainty.

With regards to human vision, the spectral features of each subject in the visible region can be described using the CIE colour space coordinates, commonly referred to as CIELAB or CIE $L^*a^*b^*$. The coordinate L^* represents lightness of colour where L^*=0 indicates black and L^*=100 indicates diffuse white. The coordinates a^* and b^* represent colour along the magenta-green and yellow-blue continuums, respectively. The resulting CIE $L^*a^*b^*$ values for all subjects are plotted on the left side in Figure 3.

Cooksey, Catherine; Allen, David; Tsai, Benjamin.
”Reference data set and variability study for human skin reflectance.”
Paper presented at 29th Quadrennial Session of the International Commission on Illumination (CIE), Washington, D.C., United States. June 14, 2019 - June 22, 2019.

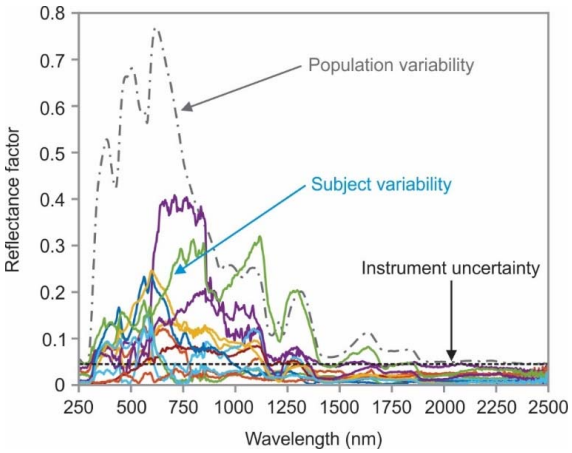


Figure 2 – The instrument uncertainty (black dashed), subject variability for several subjects (solid coloured), and population variability for all subjects (grey dot-dashed).

Additionally, the Euclidean distance between the CIE *L*a*b** coordinates for each subject with respect to the coordinates for the representative of the mean was calculated according to:

$$\Delta E^*_{ab} = \sqrt{(L^*_i - L^*_m)^2 + (a^*_i - a^*_m)^2 + (b^*_i - b^*_m)^2} \quad (2)$$

where

L^*_i, a^*_i, b^*_i are the CIE *L*a*b** coordinates for subject *i*;
L^*_m, a^*_m, b^*_m are the CIE *L*a*b** coordinates for the representative of the mean.

The ΔE^*_{ab} for each subject are plotted as a histogram on the right side of Figure 3.

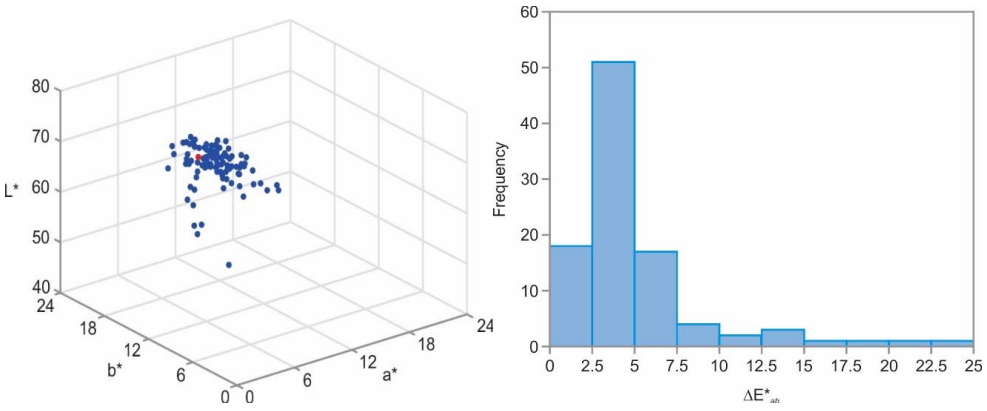


Figure 3 – Left: A three-dimensional plot of the CIE *L*a*b values for all subjects. The CIE *L*a*b** coordinates denoted in red are those of the representative of the mean. Right: A histogram of ΔE^*_{ab} for each subject with respect to the representative of the mean.**

4 Conclusions

The reference data set presented here provides reliable, traceable spectra of human skin reflectance with stated measurement uncertainties. The data set reveals a range and distribution of "typical" human skin reflectance values. While a larger sample size would better reflect the population at large, it can be expected that the variability of human skin reflectance is no smaller than the distribution represented by this set. Indeed, the range of colour

Cooksey, Catherine; Allen, David; Tsai, Benjamin.
"Reference data set and variability study for human skin reflectance."
Paper presented at 29th Quadrennial Session of the International Commission on Illumination (CIE), Washington, D.C., United States. June 14, 2019 - June 22, 2019.

coordinates calculated for this data set is similar to that of the database, which includes nearly 1000 skin colour measurements (Xiao, 2016).

References

COOKSEY, C.C., ALLEN, D.W. 2013. Reflectance Measurements of Human Skin from the Ultraviolet to the Shortwave Infrared (250 nm to 2500 nm). *Proc. SPIE*, 8734, 87340N.

COOKSEY, C.C., TSAI, B.K., ALLEN, D.W. 2014. A Collection and Statistical Analysis of Skin Reflectance Signatures for Inherent Variability over the 250 nm to 2500 nm Spectral Range. *Proc. SPIE*. 9082, 908206.

COOKSEY, C.C., TSAI, B.K., ALLEN, D.W. 2015. Spectral Reflectance Variability of Skin and Attributing Factors. *Proc. SPIE*. 9461, 94611M.

COOKSEY, C.C., ALLEN, D.W., TSAI, B.K. 2017. Reference Data Set of Human Skin Reflectance. *J. Res. NIST*, 122, https://doi.org/10.6028/jres.122.026.

NIST 1994. *NIST Technical Note 1297: Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. Washington, DC: US Department of Commerce.

NIST 1998. *NIST Special Publication 250-48: Spectral Reflectance.* Washington, DC: US Department of Commerce.

VENABLE, W.H., HSIA, J.J., WEIDNER, V.R. 1977. Establishing a Scale of Directional-hemispherical Reflectance Factor I: The Van den Akker Method. J Res. NBS, 82, 29-55.

XIAO, K., YATES, J.M., ZARDAWI, F., SUEEPRASAN, S., LIAO, N., GILL, L., LI, C., WUERGER, S. 2016. Characterising the Variations in Ethnic Skin Colours: a New Calibrated Data Base for Human Skin. *Skin Res. and Technology*, 0, 1.

Cooksey, Catherine; Allen, David; Tsai, Benjamin.
"Reference data set and variability study for human skin reflectance."
Paper presented at 29th Quadrennial Session of the International Commission on Illumination (CIE), Washington, D.C., United States. June 14, 2019 - June 22, 2019.

Supplementing rigorous electromagnetic modeling with atomistic simulations for optics-based metrology

Bryan M. Barnes*, Hui Zhou, Richard M. Silver, and Mark-Alexander Henn,

Nanoscale Device Characterization Division, Physical Measurement Laboratory,
National Institute of Standards and Technology, 100 Bureau Drive MS 8423,
Gaithersburg, MD, USA 20899-8423

ABSTRACT

The successful combination of electromagnetic scattering simulations and optical measurements allows for the quantification of deep-subwavelength features, including thicknesses via ellipsometry and parameterized geometries via scatterometry. Although feature size reduction has slowed in recent years, nanoelectronics still yields ever-smaller structures, thus optical measurement capabilities are ever-challenged. The critical problem is that the optical properties of materials often become thickness dependent at sub-5 nm, greatly complicating accurate fitting. These optical properties can be characterized empirically using ellipsometry and used with other prior information to reduce uncertainties via hybrid metrology, but atomistic modeling offers a unique perspective on the macroscopic optical response from features with dimensions only a few atoms in width. To illustrate the potential of such modeling, we have performed a series of density-functional theory (DFT) calculations for an ultrathin film, Si with hydrogen-terminated Si(111) surfaces. Kohn-Sham wavefunctions determined in DFT are instrumental in solving for the dielectric tensor of these configurations, as the in-plane and out-of-plane components can differ greatly with respect to incident wavelength and Si thickness. Techniques for DFT and dielectric tensor determination are reviewed, highlighting both their limitations and potential for improving optics-based metrology. The thickness- and wavelength-dependence of the resulting tensor components are parameterized using Tauc-Lorentz and Lorentz oscillators. Using an illustration from goniometric reflectometry, the quantitative effects upon dimensional metrology of employing the full thickness-dependent dielectric tensor are compared against simpler approximations of these optical properties. Reductions in parametric uncertainty in the thickness and optical constants are evaluated with a prior knowledge of the ultrathin film's thickness with uncertainties.

Keywords: optical metrology, dielectric tensor, atomistic simulation, density function theory, goniometric reflectometry

1. INTRODUCTION

The challenge of monitoring the smallest features, or critical dimensions (CDs) in the fabrication of present-day silicon-based semiconductors is already a daunting task due to the vast numbers of devices that are patterned across a 300 mm diameter wafer and the limited time allowed for these measurements [1]. While scanning probe techniques such as scanning electron microscopy (SEM) are well-suited for localized measurements of CDs at current sub-10 nm CD dimensions, optics-based measurements are equally critical for CD process control as well as for the inspection of wafers for misalignment between subsequent patterning layers (overlay metrology) and patterning errors (defect metrology) [2]. This task has continuously been made more difficult by the constant downscaling of CDs required to increase transistor density, evidenced by the success of Moore's Law over recent decades. While the rate of transistor density increase may have slowed recently [3], novel approaches are being sought not just to reduce CDs with current materials but also to add new materials that would improve computational capabilities.

Both two-dimensional (2D) materials and ultrathin films (defined here as thickness $d \leq 5$ nm) are emerging, technologically relevant candidates that might be incorporated into nanoelectronics. Often, metrology methods must be adapted or improved to measure these small dimensions. For example, SEM dimensional measurements for 10 nm to 12 nm features were improved through comparing experimental line scans against simulations

* bryan.barnes@nist.gov

Barnes, Bryan; Zhou, Hui; Silver, Richard; Henn, Mark Alexander.
"Supplementing rigorous electromagnetic modeling with atomistic simulations for optics-based metrology."
Paper presented at SPIE Optical Metrology 2019, Munich, Germany. June 24, 2019 - June 26, 2019.

modeled using the underlying physics of the scattering of incident electrons and of secondary electron emission [4]. Similarly, to reduce potential sources of error, model-based optical measurement techniques must faithfully incorporate the fundamental equations of electromagnetism, Maxwell's equations, into the modeling especially for an ultrathin medium or a 2D material. Specifically, one should investigate if the models that have worked well for thicker or wider features continue to do so at low dimensions, or if there exists a potential for a "break down" or practical failure.

Here, let us assume that the free charge $\rho \equiv 0$ in non-magnetic media thus, $\nabla \cdot \boldsymbol{\epsilon}\mathbf{E} = 0$ where $\boldsymbol{\epsilon}$ is the dielectric function, which is a function of the energy (or wavelength) of the incident light. There are two relevant known phenomena that should be considered for 2D and ultrathin films: first, the possible thickness-dependence of $\boldsymbol{\epsilon}$, and second, the potential anisotropy in $\boldsymbol{\epsilon}$ due to the near-planar or planar geometry of the ultrathin or 2D materials, respectively. Already, it has been well-established experimentally that thickness dependence exists for silicon-on-insulator films and a thickness-dependent energy shift has been reported in key peaks in $\boldsymbol{\epsilon}$, with quantum confinement speculated or identified as the root cause [5]. Furthermore, 2D and ultrathin films may also be anisotropic in their macroscopic optical response. Note, however, that in practice this anisotropy may not be observable due to the noise in the measurement. The resulting optical response which may be minimally anisotropic is therefore often treated as isotropic.

A thorough understanding of the anisotropy and thickness-dependence of $\boldsymbol{\epsilon}$ may be critical not just for the measurement of individual films but also to the role of measurements (e.g., spectroscopic ellipsometry) as input for the scatterometric fitting of nanoelectronic devices. For example, thin film characterization has already been carried out using spectroscopic ellipsometry and the measured dielectric function has been used as input for the scatterometic fitting of devices' dimensions and optial properties [6]. Model-based scatteromety with its parametric values and uncertainties may be impacted by the underlying anisotropy or thickness dependence of $\boldsymbol{\epsilon}$. Challenges in measuring ultrathin films may propagate into scatterometric or other optics-based determinations of dimensions.

This paper examines through simulations potential challenges for the measurement of ultrathin silicon films with thickness below 5 nm and suggests initial approaches for addressing these issues. Here, the electronic states of ultrathin Si featuring (111) surfaces covered with hydrogen (H-terminated Si(111)) using atomistic modeling are calculated, a computationally well-studied materials system (e.g., [7–9]). Note, the hydrogen is added to facilitate the atomistic calculations by reducing the surface free energy, but in practice H-terminated Si(111) is realizable only in vacuum. The expected macroscopic optical properties are assessed from these electronic states for seven thicknesses of H-terminated Si(111) with the Si thickness d_{Si} ranging from 0.7 nm to 6.4 nm. Here, the Random Phase Approximation (RPA) is used to determine the dielectric tensor defined as

$$\boldsymbol{\epsilon} = \begin{bmatrix} \tilde{\epsilon}_{xx} & 0 & 0 \\ 0 & \tilde{\epsilon}_{yy} & 0 \\ 0 & 0 & \tilde{\epsilon}_{zz} \end{bmatrix}, \tag{1}$$

where $\tilde{\epsilon} = \epsilon_1 + i\,\epsilon_2$. For these ultrathin films, the z-axis shall be defined as normal to the xy plane (the sample plane in Fig. 1), and $\tilde{\epsilon}_{xx} = \tilde{\epsilon}_{yy}$. Parametrization and curve fitting of ϵ_{xx} and ϵ_{zz} yield functional dependences of $\tilde{\epsilon}(d_{\mathsf{Si}})$, to allow the evaluation of its spectroscopic properties and to compare against previous calculations. While spectroscopic ellipsometry is expected to remain the dominant optical technology in industry for measuring CDs, we concentrate this work on evaluations of goniometric reflectometry. The "ground truth" for our illustration will be $\tilde{\epsilon}(d_{\mathsf{Si}})$, and the illustration will demonstrate the "experimental" determination of d_{Si}, n, and k.

Goniometric reflectometry has been integral to our group's dimensional measurements using scatterfield microscopy, illustrated schematically as Fig. 1. By employing Köhler illumination and by forming a conjugate to the back focal plane of the objective lens (CBFP), the angles of incidence at the sample can be manipulated at the CBFP. Not only have annular, dipole, and quadrupole illumination been realized in our laboratories [10, 11], but also angular reflectometry scans [12–14]. As an aperture is shifted in the CBFP, the illumination angle θ_i is well-defined but the incident beam has a finite width; decreasing the aperture diameter at the CBFP improves angular resolution at a cost of reduced incident intensity. Assuming the sample yields no higher-order scattering, the reflectance angle θ_r directs the light to the objective within its numerical aperture. Use of a reference sample

Barnes, Bryan; Zhou, Hui; Silver, Richard; Henn, Mark Alexander.
"Supplementing rigorous electromagnetic modeling with atomistic simulations for optics-based metrology."
Paper presented at SPIE Optical Metrology 2019, Munich, Germany. June 24, 2019 - June 26, 2019.

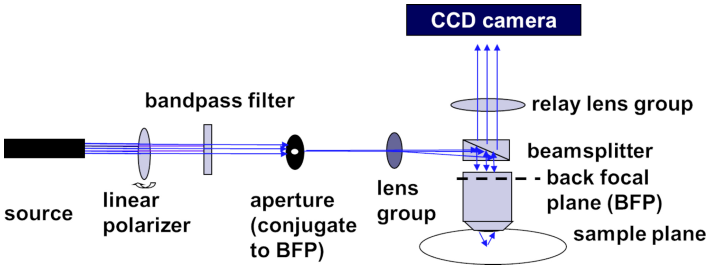


Figure 1. Schematic of angle-resolved scatterfield microscopy for goniometric reflectivity measurements. After Ref. [12].

can enable conversion of measured intensities into reflectivity values while accounting for the physical optics [15].

2. CALCULATING THE DIELECTRIC TENSOR FOR AN ULTRATHIN FILM

The determination of optical constants from *a priori* information has been a key objective for several decades [16]. One well-known candidate method has been Density-Functional Theory (DFT) [17]. DFT allows the quantum mechanical computation of the ground-state electronic structure of many-body systems from first principles. DFT is not the only available method; some researchers have shown improved results using tight binding calculations [18]. This ground-state electronic structure can then be used to calculate the imaginary part ϵ_2 of the diagonal elements of the dielectric tensor as a function of energy. To enable model-based optical metrology at multiple wavelengths λ, the functional form of the key dielectric tensor elements have been fitted parametrically. The advantages and shortcomings of these approaches are discussed.

2.1 Density-Functional Theory

DFT allows the quantum mechanical computation of the ground-state electronic structure of many-body systems from first principles. Specifically,

$$\widehat{H}\Psi_j(x_1,\ldots,x_N) = \mathrm{E}_j\Psi_j(x_1,\ldots,x_N), \tag{2}$$

where Ψ_j is an N-electron wavefunction that is the j-th eigenstate of the Hamiltonian $\widehat{H}$ with energy eigenvalue E_j, is not directly solvable but by using the Kohn-Sham (KS) equations [19], this many-body problem is reduced to a series of single-particle Schrödinger equations. The ground state energy of the interacting system is found by solving the single-particle Schrödinger equation using a local (but fictitious) external potential V_s that is a functional of the electron density ρ such that

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_s[\rho](\boldsymbol{r})\right]\phi_{\mathsf{KS}_j} = \varepsilon_j\phi_{\mathsf{KS}_j}(\boldsymbol{r}), \tag{3}$$

where ε_j is the j-th KS eigenenergy and ϕ_{KS_j} the j-th KS wavefunction, and summing the squares of the lowest N occupied orbitals to determine the electron density

$$\rho_0(\boldsymbol{r}) = \sum_{j=1}^{N}|\phi_{\mathsf{KS}_j}|^2. \tag{4}$$

It is fairly common to approximate Ψ_j using ϕ_{KS_j} and E_j using ε_j, sometimes with some empirically based rescaling of the epsilons.

Here, DFT yields estimations of the KS electron wavefunctions for the various thicknesses d_{Si} of ultrathin H-terminated Si(111), with one layer shown as Fig. 2; these KS wavefunctions are integral to the *ab initio* determination of the macroscopic optical response as a function of thickness, including anisotropy [20]. Solutions

Barnes, Bryan; Zhou, Hui; Silver, Richard; Henn, Mark Alexander.
"Supplementing rigorous electromagnetic modeling with atomistic simulations for optics-based metrology."
Paper presented at SPIE Optical Metrology 2019, Munich, Germany. June 24, 2019 - June 26, 2019.

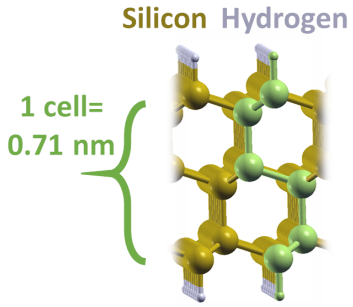


Figure 2. Representative schematic of modeled atomic structure. Atomic model shown for a single layer (i.e., a one-layer slab) of hydrogen-terminated Si(111) for density-functional theory (DFT).

have been calculated using the open-source DFT code Quantum ESPRESSO [21–23] (QE), with optimized norm-conserving Vanderbilt pseudopotentials [24, 25] "Si.upf" and "H.upf" as obtainted from [26, 27]. These DFT simulations utilized an energy cut-off at 680 eV and a 30 x 30 x 1 Monkhorst-Pack k-point sampling, but fewer k-vectors were required to compute $\boldsymbol{\epsilon}$. A notable feature of QE is the 3-D periodicity of the simulation domain. Each slab (i.e., stack of n layers where $n = 1, \ldots, 7$) has a thickness

$$z_{\mathsf{domain}} = d_{\mathsf{Si}} + d_h + d_v, \tag{5}$$

where $d_h = 0.32$ nm is the total length of the H-Si bonds on top and bottom, d_v is a sufficiently large thickness of vacuum ($d_v = 2.5$ nm) such that the top and bottom of the slab do not interact, and

$$d_{\mathsf{Si}}(n) = d_{\mathsf{Si}_1} + d_{\mathsf{Si}_n}(n-1), \tag{6}$$

where n is the number of layers, $d_{\mathsf{Si}_1} = 0.71$ nm, and $d_{\mathsf{Si}_n} = 0.94$ nm, taking into account the projection in z of the extra bond between Si layers once $n > 1$.

DFT can only approximate the actual wavefunctions. Furthermore, the quality of the DFT approximation varies with the inputs to and methodology of the calculation, such as the choice of exchange-correlation functional that uses the electron density to describe the intricate many-body effects within a single particle [17] and the corresponsing pseudopotentials tailored to that approach. The latter sidestep the need to consider the individual electrons in core orbitals. Here, the Perdew-Burke-Ernzerhof (PBE) exchange correlations have been used. Hybrid methods and pseudopotentials yield better agreement with experimental values but at great computational expense [7]. For this study, the resources required to implement a hybrid potential were deemed prohibitive given the multiple simulations required to compute the variations at each thickness, and a simpler pseudopotential was chosen. Nevertheless, these simplified DFT simulations demonstrate the *relative* changes in the wavefunctions with respect to thickness for ultrathin Si that prove critical in the calculation of $\boldsymbol{\epsilon}(d)$.

2.2 Determining Optical Constants

Having evaluated a set of the Kohn-Sham wavefunctions at each thickness, the macroscopic response to electromagnetic waves can be calculated with respect to the energy of the incident light. The open-source many-body simulation code YAMBO [28, 29] has been used to determine $\tilde{\epsilon} = \epsilon_1 + i\epsilon_2$ both in the plane of the Si slab ($\tilde{\epsilon}_{xx}$) and perpendicular to the Si(111) surfaces ($\tilde{\epsilon}_{zz}$).

As with the DFT, one can select from a range of approximations when solving for ϵ. In this work, the Independent Particle assumption of the Random Phase Approximation (RPA-IP) was utilized without including local fields, see Fig. 3. These local fields are known to play a key role in the physics of surfaces, and a strong effect of local fields is expected perpendicular to the surface plane due to the abrupt change in the electronic density. [30]. Another alternative would be the use of the Bethe-Salpeter (BSE) equations after the DFT calculations to improve the results, but this alternative is computationally expensive [31].

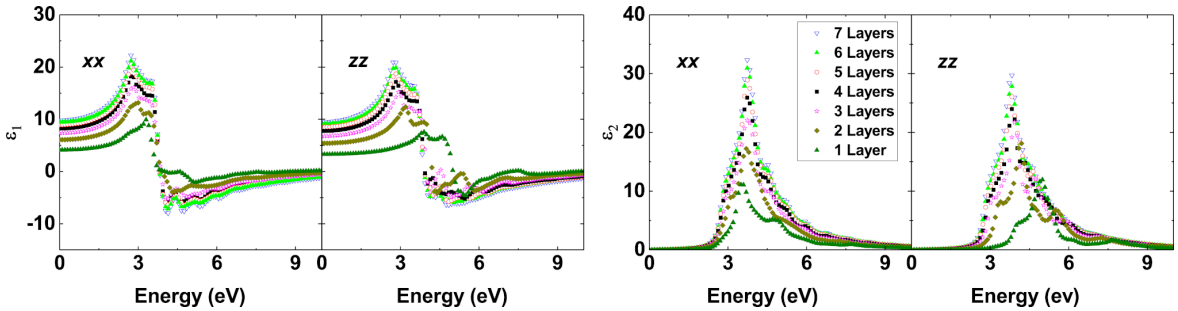


Figure 3. Calculated ϵ_1 (left) and ϵ_2 (right) for the $\tilde{\epsilon}_{xx}$ and $\tilde{\epsilon}_{zz}$ tensor components, based on DFT wavefunctions and the Random Phase Approximation in the Independent Particle approximation.

Here, the code solves for solutions to the Dyson-like equation [29]

$$\chi^0_{\mathbf{GG'}}(\mathbf{q},\omega) = \frac{2}{N_k\Omega}\sum_{nm\mathbf{k}} \rho_{nm\mathbf{k}}(\mathbf{q},\mathbf{G})\rho^*{}_{nm\mathbf{k}}(\mathbf{q},\mathbf{G'}) \times \left[\frac{f_{m\mathbf{k}}(1-f_{n\mathbf{k}-\mathbf{q}})}{\omega-(\epsilon_{m\mathbf{k}}-\epsilon_{n\mathbf{k}-\mathbf{q}})-i\eta} - \frac{f_{m\mathbf{k}}(1-f_{n\mathbf{k}-\mathbf{q}})}{\omega-(\epsilon_{m\mathbf{k}}-\epsilon_{n\mathbf{k}-\mathbf{q}})+i\eta}\right], \quad (7)$$

where χ^0 is the permittivity solved for BZ vectors $\mathbf{GG'}$, N_k is the number of reciprocal vectors $\mathbf{k}$, $\mathbf{q}$ is a reciprocal vector also, ω is the energy of the incident light, Ω is the uniot cell volume, η is an infintesimal, n, m indices represent band indexes, $f_{n\mathbf{k}}$ and $\epsilon_{n\mathbf{k}}$ are the occupations and the energies of the Kohn-Sham states, respectively, and the terms

$$\rho_{nm\mathbf{k}}(\mathbf{q},\mathbf{G}) = \langle n\mathbf{k}|e^{i(\mathbf{q}+\mathbf{G})\cdot\hat{\mathbf{r}}}|m\mathbf{k}-\mathbf{q}\rangle\,, \quad (8)$$

are oscillator strengths; further details can be found in Ref. [28].

To clearly show the aggregated effects of these approximations, Figure 4 shows comparisons of our DFT against data from the literature. In Fig. 4(a), experimental data [32] are shown for bulk Silicon. These simulated values are shifted to the left of the plot relative to the experimental data, consistent with the methods used. The simulation data would have to be shifted to larger energies, or "blue-shifted" in wavelength to attempt to match the experimental positions of these peaks. In Fig. 4(b), our simulations are compared against hybrid DFT calculations from Vazhappilly and Micha [7] (reported as the average of the ϵ_{xx} and ϵ_{zz} components) from four layers of H-terminated Si(111). These curves are also shifted in energy, with discrepancies in the magnitude. However, it is shown in Ref. [7] that such simpler pseudopotentials can yield similar values to that of hybrid potentials if the energies are shifted prior to the determination of $\boldsymbol{\epsilon}$ - this affects not only peak position but also the magnitude difference they report.

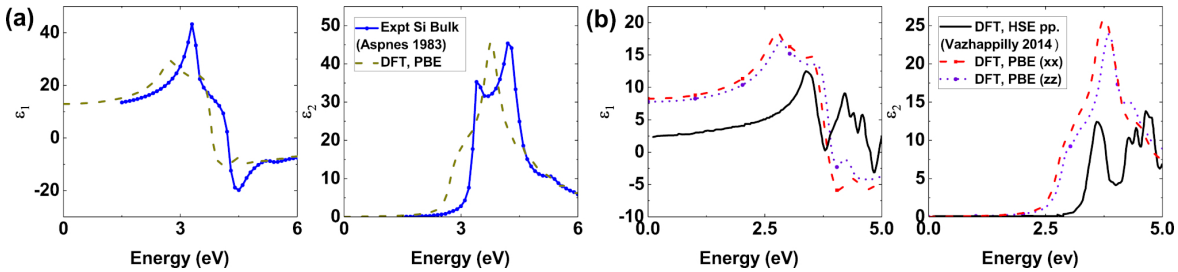


Figure 4. Effects of approximations upon calculated dielectric functions. (a) ϵ_1 and ϵ_2 DFT simulations for bulk Si using Perdew-Burke-Ernzerhof (PBE) exchange correlations and the Random Phase Approximation in the Independent Particle approximation (RPA-IP), compared to experimental values for crystalline silicon [32]. The shift of key peaks from DFT with PBE and RPA-IP is clearly observable. (b) ϵ_1 for four layers of hydrogen-terminated Si(111) as calculated here using PBE and RPA-IP compared to results from Vazhappilly and Micha [7] using a Heyd–Scuseria–Ernzerhof (HSE) hybrid exchange correlation [33].

Acknowledging these discrepancies, the dielectric tensor calculated from DFT and RPA-IP will be treated as the basis for an illustration of reflectometry from ultrathin Si(111). To demonstrate the potential for supplementing model-based parametric fitting, details regarding the thickness will be used as *a priori* information in a hybrid metrology approach [34, 35]. The ultimate goal, however would be to perform calculations with so few approximations such that after the derivation of ϵ, the results could directly supplement the optical *scatterometric* parametric fitting of complex structures containing the material of interest. This would be similar to the experimental measurements of thin films for this purpose as discussed in the Introduction.

2.3 Parameterizing Optical Constants

To define the dielectric function with respect to wavelength λ and thickness d_{Si}, a multiple-oscillator model has been used to fit ϵ_1 and ϵ_2 iteratively. Specifically, a Tauc-Lorentz oscillator has been paired with an Urbach tail and combined with three Lorentz oscillators. While several commercial software alternatives exist for such fitting, functions have been fitted in-house using analytical treatments of optical-constant models recently published by Larruquert and Rodríguez-de Marcosa [36, 37]. Using their nomenclature,

$$\tilde{\epsilon}_{\mathsf{an}}(E) = \tilde{\epsilon}_{\mathsf{U-an}}(E; A, E_0, E_g, C, a, E_c) + \tilde{\epsilon}_{\mathsf{TL-an}}(E; A, E_0, E_g, C, a, E_c) + \sum_{j=1}^{3} \left(\tilde{\epsilon}_A(E; E_j, B_j, a_j) - 1\right), \tag{9}$$

where $\tilde{\epsilon}_{\mathsf{U-an}}(E; A, E_0, E_g, C, a, E_c)$ is the analytic expression of an Urbach tail as defined in Eqns. (7), (8), and (15) of Ref. [36], $\tilde{\epsilon}_{\mathsf{TL-an}}(E; A, E_0, E_g, C, a, E_c)$ is a Tauc-Lorentz oscillator as defined in Eqns. (7), (8), (16), (17), and (18) of that same work, and $\sum_j(\tilde{\epsilon}_A(E; E_j, B_j, a_j) - 1)$ is a sum of Lorentz oscillators that has been derived from a Sellmeier model, defined as Eqn. (7) in [37]. The Tauc-Lorentz with an Urbach tail requires six parameters: energy gap E_g, central oscillator energy E_0, oscillator width C, amplitude A, non-zero parameter a, and the connection energy E_c which defines the transition from Urbach to Tauc-Lorentz. Each Lorentz oscillator has three parameters: amplitude B_j, a_j that equals each oscillator's half-width, and oscillator energy E_j. In practice, to simplify the fitting we assume $E_c = E_g + 0.1$ eV while $a \equiv 0.008$ eV, yielding a thirteen parameter fit for ϵ_{xx} and ϵ_{zz}, with best fit parameters presented in Tables 1 and results shown in Fig. 5.

Using the following conversions,

$$\lambda[\mathrm{nm}] = \frac{1239.8[\mathrm{eV}]}{\mathrm{E}}, \tag{10}$$

$$n_{xx}(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}})) = \sqrt{\frac{\sqrt{\epsilon_1(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}}))^2 + \epsilon_2(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}}))^2} + \epsilon_1(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}}))}{2}}, \tag{11}$$

$$k_{xx}(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}})) = \sqrt{\frac{\sqrt{\epsilon_1(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}}))^2 + \epsilon_2(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}}))^2} - \epsilon_1(d_{\mathsf{Si}}, \lambda, \boldsymbol{\nu}(d_{\mathsf{Si}}))}{2}}, \tag{12}$$

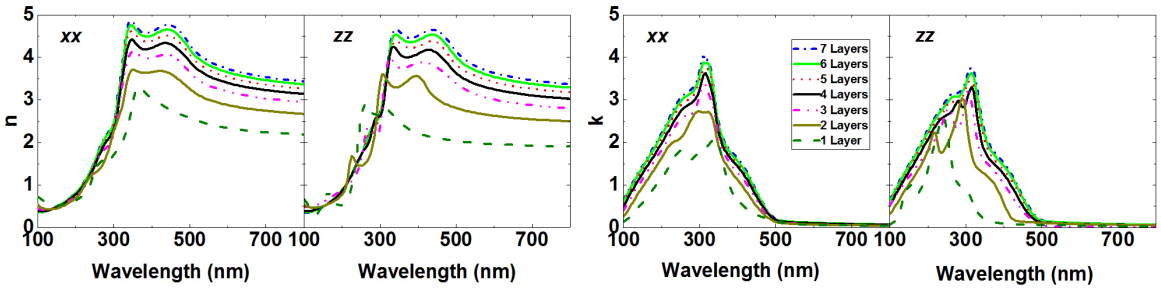


Figure 5. Plots of fitted values for n_{xx}, n_{zz}, k_{xx}, and k_{zz} as functions of wavelength and of the number of Si layers in the ultrathin film.

where E is the photon energy and λ the photon wavelength, and $\boldsymbol{\nu}(d_{\mathsf{Si}})$ the vector of parameters shown in Table 1, the optical properties n_{xx} and k_{xx} can be established in-plane. For the direction perpendicular to the surface and n_{zz}, k_{zz}, we substitute the $\boldsymbol{\nu}(d_{\mathsf{Si}})$ using the parameters shown in Table 2.

Table 1. Parametric variables for characterizing $\epsilon_{1_{xx}}(\lambda, d)$ and $\epsilon_{2_{xx}}(\lambda, d)$ using a Tauc/Lorentz function with an Urbach tail (TLU) using the formulas of [36] and three Lorentz oscillators (LO1, LO2, LO3) using [37]. For TLU, the parameter $E_c > E_g$; to ensure this, $E_c = E_g + 0.1$ eV. For these fits, $a \equiv 0.008$ eV.

Layers		TLU				LO1			LO2			LO3		
Si Layers	d_{Si} (nm)	A	E_0	E_g	C	E_1	B_1	a_1	E_2	B_2	a_2	E_3	B_3	a_3
1	0.7	53	2.22	3.28	2.76	3.53	12.8	0.23	3.97	7.6	0.40	4.85	10.0	0.40
2	1.6	179	2.38	2.87	1.29	3.69	13.6	0.22	4.09	19.3	0.41	5.25	9.7	0.60
3	2.6	308	2.42	2.68	1.07	3.70	22.5	0.26	3.93	8.1	0.18	4.63	24.2	0.64
4	3.5	288	2.38	2.77	1.13	3.70	19.3	0.21	3.90	8.5	0.16	4.47	36.3	0.77
5	4.5	357	2.41	2.70	1.03	3.74	33.5	0.25	3.98	3.2	0.16	4.56	35.7	0.75
6	5.4	384	2.41	2.69	1.01	3.75	36.9	0.25	4.00	1.7	0.09	4.54	38.8	0.75
7	6.4	391	2.40	2.71	1.03	3.75	37.3	0.24	3.99	2.0	0.06	4.53	40.0	0.73

Table 2. Parametric variables for characterizing $\epsilon_{1_{zz}}(\lambda, d)$ and $\epsilon_{2_{zz}}(\lambda, d)$ using a Tauc/Lorentz function with an Urbach tail (TLU) using the formulas of [36] and three Lorentz oscillators (LO1, LO2, LO3) using [37]. For TLU, the parameter $E_c > E_g$; to ensure this, $E_c = E_g + 0.1$ eV. For these fits, $a \equiv 0.008$.

Layers		TLU				LO1			LO2			LO3		
Si Layers	d_{Si} (nm)	A	E_0	E_g	C	E_1	B_1	a_1	E_2	B_2	a_2	E_3	B_3	a_3
1	0.7	38	3.15	4.02	0.64	4.98	41.7	0.38	7.83	13.6	0.70	0.00	0.0	0.00
2	1.6	391	2.87	3.02	0.73	4.17	19.4	0.19	4.36	18.4	1.03	5.58	9.8	0.24
3	2.6	241	2.38	3.02	1.70	3.99	8.5	0.12	3.86	7.1	0.15	4.74	13.9	0.26
4	3.5	299	2.46	2.82	1.12	3.85	24.2	0.21	4.36	6.8	0.19	4.88	36.3	0.84
5	4.5	437	2.51	2.66	0.88	3.84	41.2	0.31	4.63	17.4	0.38	5.38	19.4	0.84
6	5.4	375	2.45	2.72	0.99	3.82	38.8	0.28	4.53	19.5	0.45	5.17	24.6	0.94
7	6.4	366	2.43	2.74	1.02	3.81	40.0	0.27	4.55	27.5	0.53	5.35	18.7	0.93

The differences in the resulting n_{xx}, n_{zz}, k_{xx} and k_{zz} with varying number of layers for a fixed wavelength of $\lambda = 300$ nm are presented in Fig. 6.

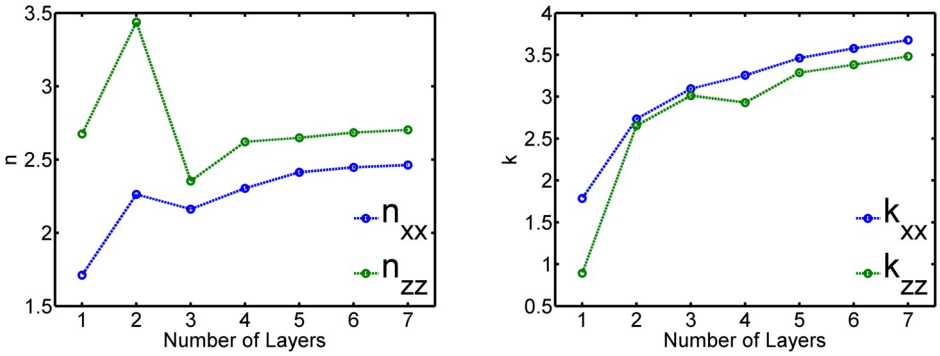


Figure 6. Difference in optical constants for different number of layers for $\lambda = 300$ nm assuming anisotropy.

3. IMPROVING THE REFLECTOMETRY FOR ULTRATHIN FILMS

As noted in the Introduction, our group has published extensively on angle-resolved scans within a high-magnification platform. A major challenge of this approach however is parametric correlations, and to address this our group and colleagues pioneered what would be known as "hybrid metrology," which incorporates prior

information obtained by other means, such as another measurement or known process, with uncertainties. In the illustration here, goniometric reflectometry will be used to demonstrate these challenges for measuring ultrathin films and then augmented using hybrid metrology.

3.1 Angle-Based Reflectometry of Anisotropic Ultrathin Films

For complicated geometries, electromagnetic simulations are often required to compute the reflectivity or high-spatial frequency scattering. Anisotropic effects in reflectivity and scattering can be calculated in rigorous electromagnetic modeling software using e.g., the finite-difference time-domain (FDTD) method or finite-element methods. However, for this present work, the essential challenges in the treatment of the film thickness d and complex optical constants $\tilde{n}(d, \lambda, \boldsymbol{\nu})$ can be illustrated using a simple geometry, shown as Fig. 7 that yields an analytically derived reflectivity that accounts for anisotropy. Eqns. 13, 14, and 15 summarize the complex Fresnel reflection coefficients for a free-standing anisotropic film in vacuum [38].

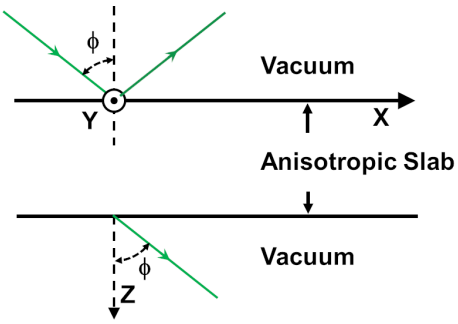


Figure 7. Polarized light incident upon an anisotropic slab from the isotropic vacuum with no substrate. Based on [38].

$$R_{pp}(\tilde{n}_{\mathrm{o1}}, \tilde{n}_{\mathrm{e1}}, d, \theta) = \frac{\frac{\tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}\cos(\theta) - A_e}{A_e + \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}\cos(\theta)} + \frac{e^{-\frac{4i\pi d_1 A_e \tilde{n}_{\mathrm{o1}}}{\lambda \tilde{n}_{\mathrm{e1}}}}(A_e - \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}|\cos(\theta)|)}{A_e + \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}|\cos(\theta)|}}{1 - \frac{e^{-\frac{4i\pi d_1 A_e \tilde{n}_{\mathrm{o1}}}{\lambda \tilde{n}_{\mathrm{e1}}}}(A_e - \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}\cos(\theta))(A_e - \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}|\cos(\theta)|)}{(A_e + \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}\cos(\theta))(A_e + \tilde{n}_{\mathrm{e1}}\tilde{n}_{\mathrm{o1}}|\cos(\theta)|)}}, \tag{13}$$

$$R_{ss}(\tilde{n}_{\mathrm{o1}}, d, \theta) = \frac{\frac{\cos(\theta) - A_o}{A_o + \cos(\theta)} + \frac{e^{-\frac{4i\pi d_1 A_o}{\lambda}}(A_o - |\cos(\theta)|)}{A_o + |\cos(\theta)|}}{1 - \frac{e^{-\frac{4i\pi d_1 A_o}{\lambda}}(A_o - \cos(\theta))(A_o - |\cos(\theta)|)}{(A_o + \cos(\theta))(A_o + |\cos(\theta)|)}}, \tag{14}$$

where

$$A_e = \sqrt{\tilde{n}_{\mathrm{e1}}^2 - \sin^2(\theta)}, \text{ and } A_o = \sqrt{\tilde{n}_{\mathrm{o1}}^2 - \sin^2(\theta)}. \tag{15}$$

The subscripts e and o denote the extraordinary and ordinary directions (i.e., zz and xx, respectively) and $\tilde{n} = n + ik$. Note, R_{ss} is independent of $\tilde{n}_{zz}$.

With these data, one may simulate the reflectivities R_p and R_s. For angle-based reflectometry, one is concerned with two incident orthogonal polarizations

$$R_p(\theta) = R_{pp}(\theta) \cdot R_{pp}(\theta)^* \tag{16}$$

$$R_s(\theta) = R_{ss}(\theta) \cdot R_{ss}(\theta)^* \tag{17}$$

where $R_p(\theta)$ and $R_s(\theta)$ are reflectivities at an incident angle θ that varies from 0 to 45 degrees, which is approximately the achievable range for our visible-light Scatterfield Microscope with an NA=0.95. The fixed incident wavelength in this example is $\lambda = 300$ nm. Examples of the simulated datasets without noise are shown in Fig. 8.

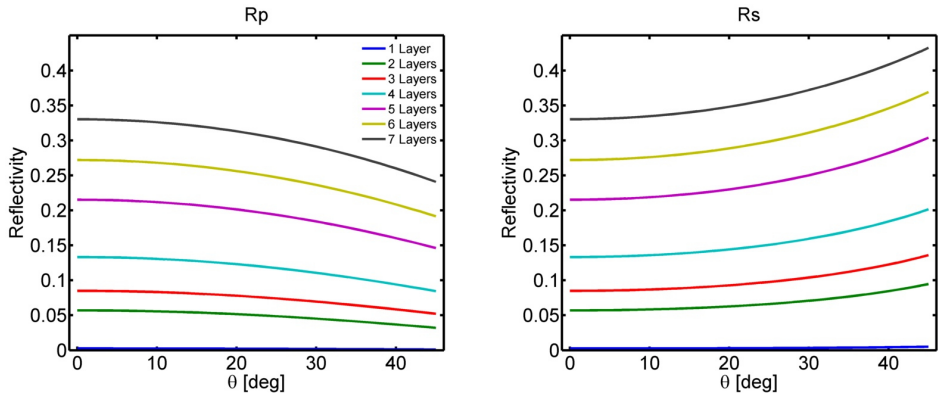


Figure 8. Example simulation data for goniometric reflectometry using the optical constants in Fig. 6 at $\lambda = 300$ nm.

3.2 Parametric Fitting

Non-linear regression is employed in our estimation of layer thickness d_{Si} and n and k values from the simulated measurement data presented in Fig. 8 with added noise. This means that based on a model that is able to approximate these data, which we will call $\mathbf{y} \in \mathbb{R}^n$, we want to determine the parameter values that best reproduce these measurements within their expected uncertainties. If we denote our model function as

$$\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^n, \ \mathbf{p} = (d_{\mathsf{Si}}, n, k) \mapsto \mathbf{f}(\mathbf{p}), \tag{18}$$

we perform the parametric fitting by minimizing

$$\chi^2(\mathbf{p}) = [\mathbf{f}(\mathbf{p}) - \mathbf{y}]^{\intercal} \boldsymbol{\Sigma}^{-1} [\mathbf{f}(\mathbf{p}) - \mathbf{y}]. \tag{19}$$

We assume that the measurement data are a noisy realization of the model, and that we have an additive error model where

$$\mathbf{y} = \mathbf{f}(\mathbf{p}) + \boldsymbol{\eta}, \tag{20}$$

the noise $\boldsymbol{\eta}$ is assumed to be normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$. In this example we assume that the errors are independent and identically distributed with zero mean and a variance of $\sigma^2 = (0.001)^2$.

The measurement data are generated based on two potential assumptions about these ultrathin films: it is either fully described by $\boldsymbol{\epsilon}_{xx}$ (the isotropic assumption) or it is described by both $\boldsymbol{\epsilon}_{xx}$ and $\boldsymbol{\epsilon}_{zz}$ (the anisotropic assumption) with parameters for their functional form given in Table 1 for the isotropic and Tables 1 and 2 for the anisotropic input, evaluated for the seven different Si thicknesses found in the same tables. Figure 9 presents the resulting estimates for the thickness $\widehat{d}_{\mathsf{Si}}$ for both the isotropic and anisotropic input data for a total of 50 realizations of noise. Both the isotropic and anisotropic input data leads to a slight systematic offset in the estimated $\widehat{d}_{\mathsf{Si}}$'s. The error bars however, cover the nominal value for each number of layers.

3.3 Supplementing Parametric Fitting using Hybrid Metrology

The most straightforward improvement to the methodology is to include prior knowledge about the layer thickness in a Bayesian sense in our regression [34, 35]. It is known that Raman spectroscopy can be used to determine Si thicknesses between 0.8 nm to 3.5 nm within a Si/SiO_2 multilayer [39] and that contrast spectroscopy, reflectometry, and Raman scattering are effective on graphene [40].

For this illustration, the effects of such prior knowledge (with uncertainty) are evaluated for the estimated parameters for several priors with fixed means and multiple variances that depend on the nominal thicknesses, such that

$$\mu_d = d_0, \ \sigma_d^2(j) = \left(\frac{d_0}{2^j}\right)^2, \ j = 1, 2, \ldots, 8. \tag{21}$$

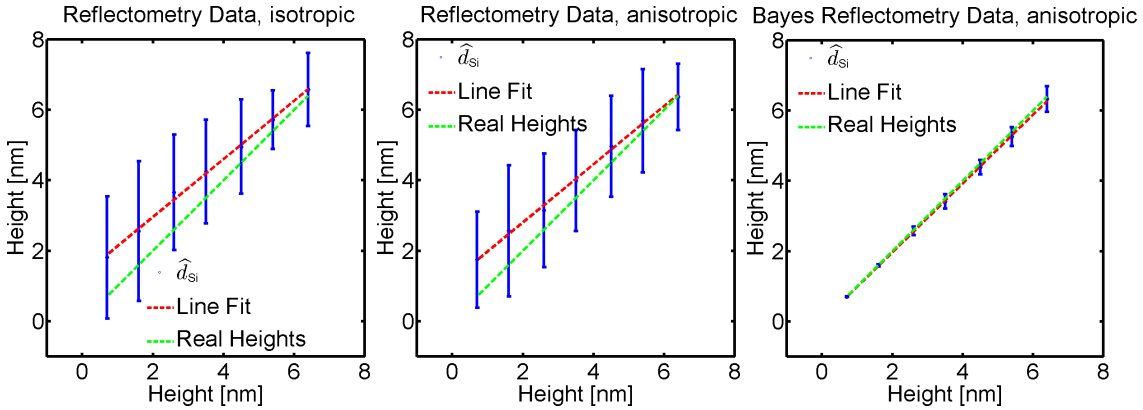


Figure 9. Nominal vs. estimated layer thickness for isotropic (left) and anisotropic (middle) input data. Nominal vs. estimated layer thickness for anisotropic input data using prior knowledge on d (right). Error bars show the parametric uncertainties (coverage factor k=1) of the height, based on 50 independent optimizations.

Changes in the parameter estimates due to the changing priors for input data for seven Si layers is presented in Fig. 10. Not only does the uncertainty in $\widehat{d}_{\mathsf{Si}}$ reduce due to the prior knowledge but due to the high correlation between d_{Si} and k, the prior knowledge of d_{Si} also helps to decrease the uncertainty in k. This is true for all numbers of layers. Appreciable changes in the parametric uncertainties in k is realized rapidly at seven layers if the uncertainty in d is less than 10 %.

The resulting thicknesses and their uncertainties for all numbers of layers with $j = 3$, i.e., $\sigma_d^2 = \left(\frac{d_0}{8}\right)^2$ is presented in the right panel of Fig. 9. The nominal and estimated layer thicknesses match very well, removing the systematic bias.

Incorporating prior knowledge on the optical constants obtained by the DFT is not as straightforward partly due to the fact that n and k change with layer thickness d_{Si}. Instead of providing a simple prior distribution for n or k, we would need to provide their probability distributions conditioned on d_{Si}. Further research is needed to achieve this task and to determine its potential effect.

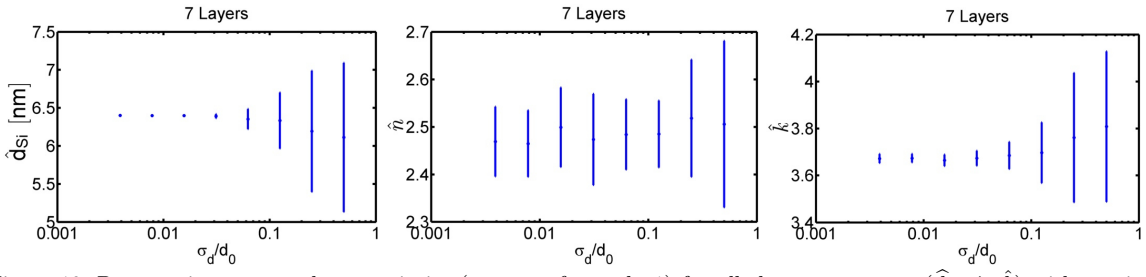


Figure 10. Parametric means and uncertainties (coverage factor k=1) for all three parameters ($\widehat{d}_{\mathsf{Si}}$, $\hat{n}$, $\hat{k}$) with varying prior knowledge of the uncertainty σ_d of the known thickness d_0.

4. CONCLUSION

Potential directions for improving optics-based metrology for ultrathin films and potentially even 2D materials have been explored using an illustration from goniometric reflectometry. The thickness-dependent optical constants of the H-terminated ultrathin Si films with (111) surfaces have been computed using the random phase approximation for independent particles and ground-state Kohn-Sham wavefunctions from density-functional theory. These atomistic calculations have yielded an understanding of the dielectric tensor $\boldsymbol{\epsilon}(d)$ and its in-plane $\tilde{\epsilon}_{xx}$ and out-of-plane $\tilde{\epsilon}_{zz}$ components, even given their approximate nature. As expected, the correlations among

Barnes, Bryan; Zhou, Hui; Silver, Richard; Henn, Mark Alexander.
"Supplementing rigorous electromagnetic modeling with atomistic simulations for optics-based metrology."
Paper presented at SPIE Optical Metrology 2019, Munich, Germany. June 24, 2019 - June 26, 2019.

parameters, especially between d and k, led to large parametric uncertainties in fits to "experimental" data generated from the calculated $\boldsymbol{\epsilon}(d)$, with increasing systematic bias as the thickness decreased. This thickness dependent effect proved dominant in our example compared to the difference in anisotropy.

Additionally, the effects of prior knowledge, particularly the mean and uncertainty in d, on all three model parameters have been studied. If d is known to within 10 % or better for a seven-layer Si slab, the decrease in the uncertainty in k is appreciable. Future work is to be performed to allow the optical constants determined from DFT to supplement the fitting, independent of or in conjunction with prior knowledge of d.

Spectroscopic ellipsometry, with some effort, may also benefit from supplemental information from atomistic modeling, but the path is less straightforward. Employing a spectroscopic approach requires determining the n and k values for a range of wavelengths, potentially adding two parameters per wavelength. One can however, reduce the complexity of this task by making use of the fitting parameters for ϵ_{xx} and ϵ_{zz} for Eqs. 11 and 12. The number of these parameters needed to determine $n(\lambda)$ and $k(\lambda)$ will change depending on the model used. Hence in this work where a Tauc-Lorentz oscillator with three Lorentz oscillators was employed, the total number of parameters is 14, $(d_{\mathsf{Si}}, A, E_o, E_g, \ldots, a_3)$ see Table 1, but can be reduced when assuming a different model, e.g., by reducing the number of Lorentz oscillators. Especially in the case where we assume the maximum number of parameters the aforementioned ansatz of incorporating prior knowledge on the optical constants should prove helpful as a way to regularize this regression problem and decrease the resulting parametric uncertainties.

References

[1] Orji, N. G., Badaroglu, M., Barnes, B. M., Beitia, C., Bunday, B. D., Celano, U., Kline, R. J., Neisser, M., Obeng, Y., and Vladár, A. E., "Metrology for the next generation of semiconductor devices," Nature Electronics 1(10), 532–547 (2018).

[2] Ma, Z. and Seiler, D. G., [Metrology and Diagnostic Techniques for Nanoelectronics], Pan Stanford (2017).

[3] Khan, H. N., Hounshell, D. A., and Fuchs, E. R., "Science and research policy at the end of Moore's law," Nature Electronics 1(1), 14 (2018).

[4] Villarrubia, J. S., Vladár, A. E., Ming, B., Kline, R. J., Sunday, D. F., Chawla, J., and List, S., "Scanning electron microscope measurement of width and shape of 10nm patterned lines using a JMONSEL-modeled library," Ultramicroscopy 154, 15–28 (2015).

[5] Price, J. and Diebold, A. C., "Spectroscopic ellipsometry characterization of ultrathin silicon-on-insulator films," J. Vac. Sci. Technol. B 24(4), 2156–2159 (2006).

[6] Ebersbach, P., Urbanowicz, A. M., Likhachev, D., and Hartig, C., "Monitoring of ion implantation in microelectronics production environment using multi-channel reflectometry," Proc. SPIE 9778, 977812 (2016).

[7] Vazhappilly, T. and Micha, D. A., "Computational modeling of the dielectric function of silicon slabs with varying thickness," The Journal of Physical Chemistry C 118(8), 4429–4436 (2014).

[8] Nakamura, J., Wakui, S., and Natoriand, A., "First-principles evaluations of dielectric constants," in [2006 International Workshop on Nano CMOS], 236–249 (2006).

[9] Agrawal, B. K. and Agrawal, S., "First-principles study of one-dimensional quantum-confined H-passivated ultrathin Si films," Appl. Phys. Lett. 77(19), 3039–3041 (2000).

[10] Sohn, Y. J., Quintanilha, R., Barnes, B. M., and Silver, R. M., "193 nm angle-resolved scatterfield microscope for semiconductor metrology," Proc. SPIE 7405, 74050R (2009).

[11] Barnes, B. M., Sohn, M. Y., Goasmat, F., Zhou, H., Vladár, A. E., Silver, R. M., and Arceo, A., "Three-dimensional deep sub-wavelength defect detection using $\lambda = 193$ nm optical microscopy," Opt. Express 21(22), 26219–26226 (2013).

[12] Silver, R., Barnes, B., Attota, R., Jun, J., Filliben, J., Soto, J., Stocker, M., Lipscomb, P., Marx, E., Patrick, H., et al., "The limits of image-based optical metrology," Proc. SPIE 6152, 61520Z (2006).

Barnes, Bryan; Zhou, Hui; Silver, Richard; Henn, Mark Alexander.
"Supplementing rigorous electromagnetic modeling with atomistic simulations for optics-based metrology."
Paper presented at SPIE Optical Metrology 2019, Munich, Germany. June 24, 2019 - June 26, 2019.

[13] Silver, R. M., Barnes, B. M., Attota, R., Jun, J., Stocker, M., Marx, E., and Patrick, H. J., "Scatterfield microscopy for extending the limits of image-based optical metrology," Appl. Opt. 46(20), 4248–4257 (2007).

[14] Silver, R. M., Barnes, B. M., Heckert, A., Attota, R., Dixson, R., and Jun, J., "Angle resolved optical metrology," Proc. SPIE 6922, 69221M (2008).

[15] Barnes, B. M., Attota, R., Quintanilha, R., Sohn, Y., and Silver, R. M., "Characterizing a scatterfield optical platform for semiconductor metrology," Measurement Science and Technology 22(2), 024003 (2010).

[16] Gopalan, S., Kircher, J., Andersen, O. K., Jepsen, O., Alouani, M., and Cardona, M., "Investigation of the dielectric tensor in YBa2Cu3O7," Physica C: Superconductivity 185, 1473–1474 (1991).

[17] Cohen, A. J., Mori-Sánchez, P., and Yang, W., "Challenges for density functional theory," Chemical Reviews 112(1), 289–320 (2011).

[18] Kriso, C., Triozon, F., Delerue, C., Schneider, L., Abbate, F., Nolot, E., Rideau, D., Niquet, Y.-M., Mugny, G., and Tavernier, C., "Modeled optical properties of sige and si layers compared to spectroscopic ellipsometry measurements," Solid-State Electronics 129, 93–96 (2017).

[19] Kohn, W. and Sham, L. J., "Self-consistent equations including exchange and correlation effects," Phys. Rev. 140(4A), A1133 (1965).

[20] Matthes, L., Pulci, O., and Bechstedt, F., "Influence of out-of-plane response on optical properties of two-dimensional materials: First principles approach," Phys. Rev. B 94, 205408 (2016).

[21] Giannozzi, P. et al., "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," Journal of Physics: Condensed Matter 21(39), 395502 (2009).

[22] Giannozzi, P. et al., "Advanced capabilities for materials modelling with Quantum ESPRESSO," Journal of Physics: Condensed Matter 29(46), 465901 (2017).

[23] Certain commercial materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials are necessarily the best available for the purpose.

[24] Hamann, D. R., "Optimized norm-conserving Vanderbilt pseudopotentials," Phys. Rev. B 88, 085117 (2013).

[25] Hamann, D. R., "ONCVPSP." http://www.mat-simresearch.com// (2018).

[26] http://www.pseudo-dojo.com// (2018).

[27] van Setten, M., Giantomassi, M., Bousquet, E., Verstraete, M., Hamann, D., Gonze, X., and Rignanese, G.-M., "The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table," Computer Physics Communications 226, 39 – 54 (2018).

[28] Marini, A., Hogan, C., Grüning, M., and Varsano, D., "yambo: An ab initio tool for excited state calculations," Computer Physics Communications 180(8), 1392–1403 (2009).

[29] Sangalli, D., Ferretti, A., Miranda, H., Attaccalite, C., Marri, I., Cannuccia, E., Melo, P., Marsili, M., Paleari, F., Marrazzo, A., et al., "Many-body perturbation theory calculations using the yambo code," Journal of Physics: Condensed Matter , in press (2019).

[30] Tancogne-Dejean, N., Giorgetti, C., and Véniard, V., "Optical properties of surfaces with supercell ab initio calculations: Local-field effects," Phys. Rev. B 92, 245308 (2015).

[31] Onida, G., Reining, L., and Rubio, A., "Electronic excitations: density-functional versus many-body Green's-function approaches," Rev. Mod. Phys. 74, 601–659 (2002).

[32] Aspnes, D. E. and Studna, A., "Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV," Physical Review B 27(2), 985 (1983).

[33] Heyd, J. and Scuseria, G. E., "Efficient hybrid density functional calculations in solids: Assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional," J. Chem. Phys. 121(3), 1187–1192 (2004).

[34] Henn, M.-A., Silver, R. M., Villarrubia, J. S., Zhang, N. F., Zhou, H., Barnes, B. M., Ming, B., and Vladár, A. E., "Optimizing hybrid metrology: rigorous implementation of bayesian and combined regression," Journal of Micro/Nanolithography, MEMS, and MOEMS 14(4), 044001 (2015).

[35] Zhang, N. F., Barnes, B. M., Zhou, H., Henn, M.-A., and Silver, R. M., "Combining model-based measurement results of critical dimensions from multiple tools," Measurement Science and Technology 28(6), 065002 (2017).

[36] Rodríguez-de Marcos, L. V. and Larruquert, J. I., "Analytic optical-constant model derived from Tauc-Lorentz and Urbach tail," Opt. Express 24(25), 28561–28572 (2016).

[37] Larruquert, J. I. and Rodríguez-de Marcos, L. V., "Procedure to convert optical-constant models into analytic," Thin Solid Films 664, 52–59 (2018).

[38] Bashara, N. and Azzam, R., [Ellipsometry and polarized light], NH, Amsterdam (1977).

[39] Khriachtchev, L., Räsänen, M., Novikov, S., Kilpelä, O., and Sinkkonen, J., "Raman scattering from very thin Si layers of Si/SiO2 superlattices: Experimental evidence of structural modification in the 0.8-3.5 nm thickness region," J. Appl. Phys. 86(10), 5601–5608 (1999).

[40] Ni, Z. H., Wang, H. M., Kasim, J., Fan, H. M., Yu, T., Wu, Y. H., Feng, Y. P., and Shen, Z. X., "Graphene thickness determination using reflection and contrast spectroscopy," Nano Letters 7(9), 2758–2763 (2007).

Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack

A. N. Johnson[1], I. I. Shinder[1], J. B. Filla[1], J. T. Boyd[1], R. Bryant[1], M. R. Moldover[1], T. D. Martz[2]
[1]NIST 100 Bureau Drive, Gaithersburg, MD, USA
[2]EPRI, 3420 Hillview Ave., Palo Alto, CA 94304
(corresponding author): Aaron.Johhson@nist.gov

Exhaust flows from coal-fired stacks are determined by measuring the flue gas velocity at prescribed points in the stack cross section. During the last 30+ years these velocity measurements have been made predominantly using S-type pitot probes. These probes are robust and inexpensive; however, S-probes measure only two components of the velocity vector and can give biased results if the stack flow has significant yaw and pitch angles. Furthermore, S-probe measurements are time intensive, requiring probe rotation (or nulling) at each traverse point to find the yaw angle. The only EPA-sanctioned alternatives to the S-probe are 5-hole probes (i.e., the prism probe and spherical probe) that also require yaw-nulling. We developed a non-nulling technique applicable to the spherical probe and two custom designed 5-hole probes that reduce testing time and may improve measurement accuracy. The non-nulling technique measures all 3 components of velocity without rotating the probe. We assessed the performance of these 5-hole probes in a coal-fired stack at the high-load (16 m/s) and the low-load (7 m/s). For the spherical probes, the non-nulling results and the nulling results were in excellent mutual agreement (< 0.1 %). For the custom probes, the non-nulling and nulling results were inconsistent: the differences were 5% at the high load and 10 % at the low load. We speculate that the nulling data for the custom probes were flawed because the non-nulling data for all the probes accurately determined the yaw and pitch angles at high and low loads. Our results demonstrate that the non-nulling technique can accurately measure flue gas flows in a coal-fired stack.

Introduction

This Introduction briefly reviews (1) the need for accurately measuring flue gas flows, (2) the current "nulling" technique for flue gas measurements and its problems, (3) the possibility of improved measurements using a non-nulling technique, and (4) the encouraging results from preliminary field tests of a non-nulling technique.

(1) The combustion gases from coal-fired power plants (CFPPs) are exhausted into large diameter (> 5 m), vertically oriented smokestacks, which emit pollutants into the atmosphere. To quantify the amount of pollutants released into the atmosphere, the total flow in these stacks must be accurately measured; however, accurate stack flow measurements are difficult. Stacks are fed by a network of elbows, reducers, fans, *etc*. which generate complex, difficult-to-measure flows. The flue gas itself causes additional difficulties because it can be hot (as high as 120°C), acidic, asphyxiating, and in some cases saturated with water vapor. Nevertheless, accurate measurements of the total flue gas flow are essential to monitor emissions of greenhouse gases (GHGs) and other hazardous pollutants.

Pollutant emissions from CFPPs are quantified by continuous emission monitoring systems (CEMS) permanently installed in the stacks. The CEMS equipment measures the concentration of each regulated pollutant as well as the total flow. Federal regulations require annual calibration of the CEMS flow monitors and concentration equipment. These calibrations are performed using an EPA protocol called a relative accuracy test audit (RATA). The flow portion of the calibration is herein referred to as the flow RATA.

(2) How Stack Flows are Currently Measured
The flow RATA maps the axial stack velocity measured along 2 orthogonal chords in the stack

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

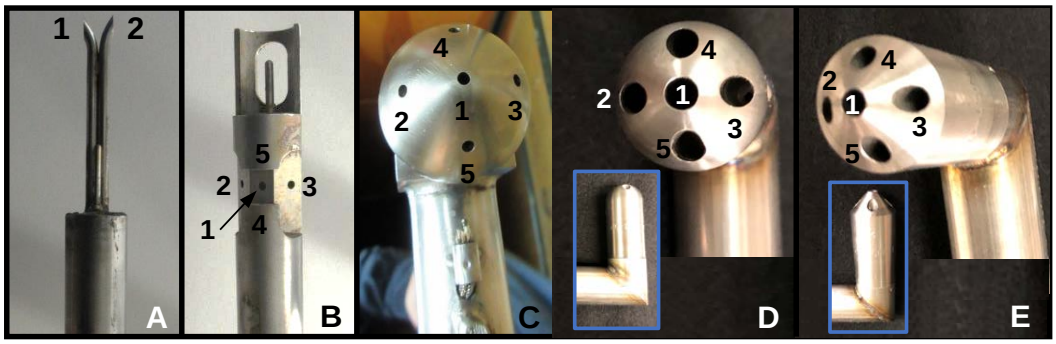


Figure 1. Pictures A, B, and C show the 3 EPA-sanctioned RATA probes including A) the S-probe, B) Prism probe, and C) Spherical Probe. The hemispherical and conical probes shown in D) and E), respectively are custom-made probes designed for non-nulling.

cross-section. A pitot probe is inserted into the flow through ports on the stack wall. On each chord, measurements are made at discrete points located at the centroids of equal area [1]. The discrete velocity measurements are integrated to determine the flow velocity, which in turn, is used to determine the correction factor for the CEMS flow meter.

Figures 1A, 1B, and 1C show the 3 RATA probe types sanctioned by the EPA including A) the S-probe, B) the prism probe, and C) the spherical probe. All 3 probes use the same measurement principle. The axial velocity is correlated to differential pressure measurements made across the probe's pressure ports. Both the prism and the spherical probe have 5 pressure ports and both measure the entire velocity vector including the pitch, yaw, and axial velocity components. In contrast, the S-probe measures only the yaw and axial velocities, and has been shown to give flow velocities that are biased high if significant pitch and yaw are present in the flow [2].

Nulling Method

The 3 EPA-sanctioned probes use a yaw-nulling method [3, 4, 5] to determine the angle of off-axis flow in the yaw direction, which we call the yaw-null angle (β_{null}). At each point on the RATA map, the probe is *nulled* by rotating it about its axis until the vector sum of the yaw and axial velocities align with pressure port 1. For a 3-D probe the nulling procedure can be accomplished in a single rotation. The S-probe requires 2 rotations. First the S-probe is nulled by rotating it about its axis until $P_{12} = 0$. A second 90° rotation orients port 1 so that it faces into the flow. Once the probe is nullled, the probe calibration parameters are used to determine the dynamic pressure (P_{dyn}), and for 3-D probes the pitch angle (α).

Errors Due to Imperfect Nulling

If the null condition is not satisfied, significant flow measurement errors can occur. The nulling errors increase with the ratio $\Delta P_{null}/P_{dyn}$, where the null-differential-pressure $\Delta P_{null} = P_{23}$ for 3-D probes and $\Delta P_{null} = P_{12}$ for the S-probe. The errors become significant when $\Delta P_{null}/P_{dyn}$ is not small relative to unity [6]. In most cases the nulling procedure is performed manually. A person rotates the probe while reading a differential pressure gauge to determine the exact yaw angle for which $\Delta P_{null} = 0$. However, transients in stack flows, noisy pressure signals, and human errors make nulling imperfect and introduce unquantified bias (*e.g.* high for an S-probe) into the measurement process.

When the velocity field has a significant yaw component, nulling the probe can be

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

time-intensive and, consequently, expensive. When mapping the flow field, several iterations are generally required to find the yaw-null angle at each traverse point. The nulling time increases in wet stacks since stack testers must frequently interrupt the measurement process to purge the probe's pressure ports of droplets or particles. Because 3-D probes have more pressure ports than S-probes, and because the diameters of these ports are smaller than the ports of S-probes, 3-D probes are more susceptible to plugging. Consequently, 3-D probes generally require more time than the S-probe to complete the flow RATA. Historically the stack flow measurement community has opted to use the more robust and economical, but less accurate, S-probe.

(3) Non-Nulling Method

Non-nulling methods determine the axial velocity without rotating the probe at each traverse point to find β_{null}. Instead, the axial velocity, the pitch angle, and the yaw-null angle β_{null} are determined with the probe oriented at zero yaw angle (*i.e.*, $\beta = 0°$ such that port 1 on the probe is aligned with the stack axis). Compared with nulling methods, non-nulling methods reduce the time needed to perform flow RATAs. CFFPs are concerned about the duration of flow RATAs because they must maintain loads stipulated by the RATA instead of loads dictated by customer supply and demand.

The non-nulling method also has the potential to improve flow measurement accuracy compared with nulling methods. First, the S-probe measures only 2 components of the velocity vector while the non-nulling method applies to 3-D probes and thereby measures the entire velocity vector. Second, Method 2F [5], which is the EPA nulling method for 3-D probes, does not address errors resulting from imperfect nulling, as discussed above.

In this manuscript we demonstrate the feasibility of accurately determining the total flow in a CFPP stack using a non-nulling method and commercially available flow RATA equipment. In previous work, we achieved 1 % accuracy when we performed flow RATAs in NIST's Scale-Model Smokestack Simulator (SMSS) using a spherical probe, even with highly-distorted flows [7, 8]. However, the SMSS facility uses ambient air as surrogate for flue gas and performs flow RATAs under laboratory conditions using laboratory grade instrumentation.

Using NIST's wind tunnel and NIST's smokestack simulator, we developed non-nulling algorithms for the spherical probe in Fig. 1C and for the 2 custom probes shown in Figs. 1D and 1E. At NIST, we calibrated these probes using our non-nulling method and also EPA's Method 2F, and then we used these probes to measure the flow velocity in a CFPP stack.

For assessing the accuracy and limitations of the non-nulling method, we selected a CFPP stack known to have complex flows. The selected stack's RATA measurement platform was only 3.8 stack diameters ($D = 6.8$ m) downstream of a 90° elbow. Moreover, upstream of the elbow, flow from two wet scrubbers merged into a single stream. Not surprisingly, the flow at the RATA platform had significant yaw-null angles that were nearly -30° at the stack wall. The flue gas was saturated with water from the wet scrubbers. The wet, particle-laden gas frequently plugged the 3-D probes; plugging increased the duration of the tests and resulted in false high (or low) axial velocity measurements both for Method 2F and the non-nulling method. We developed a statistical method based on the noisiness of the measured pressure signals to identify data affected by plugging.

The CFPP stack was equipped with an X-pattern ultrasonic flow meter system, which was used as the CEMS flow monitor. The CFPP provided us with minute by minute CEMS flow velocity data (V_{CEMS}) for the duration of the test. On average, the stability of V_{CEMS} during the flow RATAs was better than 1.5 %. We performed a 16-point flow RATA

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

using both Method 2F and the non-nulling method. The flow RATAs were performed at two loads, a high load with a flue gas velocity of 16 m/s, and at a low load of 7 m/s.

Table 1. Normalized flow velocity (V_{RATA}/V_{CEMS}) for Method 2F (M2F) and for the non-nulling method at zero yaw angle (NN, $\beta = 0$) at a high load of 16 m/s and a low load of 7 m/s.

Probe Types	Load (Repeats)	M2F	NN ($\beta = 0$)	M2F/NN -1
Spherical Probes (SP)	High (4)[a]	0.993 (2.1 %)[b]	0.994 (0.4 %)[b]	-0.1 %[c]
Custom Probes (CP)	High (4)	1.053 (0.4 %)	0.990 (0.7 %)	+5.9 %
CP/SP-1	High	6.0 %	-0.4 %	
Spherical Probes (SP)	Low (6)	1.02 (1.3 %)	1.02 (1.7 %)	0 %
Custom Probes (CP)	Low (6)	1.108 (2.0 %)	0.997 (1.6 %)	+10 %
CP/SP-1	Low	8.6 %	-2.3 %	

a) Number of repeated RATA traverses for the same probe at the same flow

b) Standard deviation of normalized RATA velocity expressed as a percent

c) Percent difference computed using 100 (M2F/NN -1)

(4) The CFPP test results are summarized in Table 1. The tabulated RATA velocities are normalized by the CEMS velocity (V_{RATA}/V_{CEMS}) to help account for flow variations during and between measurements. The data in column "M2F" are the normalized flow velocities for Method 2F; the data in column "NN ($\beta = 0$)" are the normalized non-nulling velocities obtained with the probe at a zero yaw angle.

There are 4 primary results. First, the non-nulling method and Method 2F showed excellent agreement for the spherical probes. As indicated in the last column, the difference at high load was -0.1 % and at low load the difference was 0 %.

Second, the flow results from the non-nulling method were consistent throughout the test. The percent difference of V_{NN}/V_{CEMS} determined with the spherical probes and the custom probes was only -0.4 % at high load and -2.3 % at low load.

Third, V_{NN}/V_{CEMS} is close to unity in all cases. This agreement between V_{NN} and V_{CEMS} is better than expected. The values of V_{CEMS} are based on an earlier S-probe RATA calibration that used the conventional nulling method. Our values of V_{NN} are based on a 16-point traverse that did not account for the lower velocities near the wall. If we had accounted the lower velocities, we would have found $V_{NN} < V_{CEMS}$. Note: we measured pitch angles less than 5°, so that S-probe errors related to pitch angle are negligible in this stack.

Fourth, the results of Method 2F and the non-nulling method showed poor agreement for the custom probes: the differences are 5.9 % at high load and 10 % at low load. Presently we do not understand the cause of the differences. However, we suspect these results are erroneous for the following reasons: a) the non-nulling results were consistent for all tests and agreed with the results obtained with the EPA-sanctioned spherical probe, b) in cases where RATAs based on Method 2F disagree with the S-probe nulling method, the Method 2F results are typically lower due to inherent positive biases in the S-probe [9].

Probe Calibrations in NIST's Wind Tunnel

We calibrated all 3 probe types in NIST's wind tunnel using both Method 2F [5] and the non-nulling method. Calibrations were performed in the wind tunnel's rectangular test section (1.5 m × 1.2 m) using NIST's Laser Doppler Anemometer (LDA) working standard. The metrological traceability of the LDA working standard is documented in the following references 10, 11, and 12. We use the LDA velocity (U_{LDA}) in conjunction with air density (ρ_{AIR}) in the wind tunnel to determine the dynamic pressure, $P_{dyn} = \rho_{AIR} U_{LDA}^2/2$. The wind tunnel is equipped with an automated traversing system, which positions the pitot probes to prescribed pitch angles (α) and yaw angles (β) in the test section [13, 14]. The expanded uncertainty of wind speed is less than 1 %, and the expanded uncertainties of pitch and yaw angles are 0.5°.

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

Method 2F Probe Calibrations

We calibrated 4 spherical probes, 2 hemispherical probes, and 2 conical probes. All the probes were calibrated at 11 velocities ranging from 5 m/s to 30 m/s in steps of 2.5 m/s, and at 17 pitch angles ranging from -20° to 20° in steps of 2.5°. Thus, for each probe we measured 187 combined velocity and pitch angle set points. We used *the Curve Fit Method* [6] to determine the pitch calibration factor, F_1, and the velocity calibration, F_2, at the null condition ($P_{23} = 0$). The Curve Fit Method does not require rotating the probe to the exact position where $P_{23} = 0$; instead *the pitch pressure ratio,* P_{45}/P_{12}, and *the velocity pressure ratio*, $[P_{dyn}/P_{12}]^{1/2}$, are measured over a narrow range of yaw pressures surrounding $P_{23} = 0$. By definition, the pitch pressure ratio and the velocity pressure ratio equal the respective calibration factors, $F_1 = P_{45}/P_{12}$ and $F_2 = [P_{dyn}/P_{12}]^{1/2}$ at zero yaw pressure, $P_{23} = 0$.

The measured values of the pitch pressure ratio and the velocity pressure ratio values are fit by either a 2nd or 3rd degree polynomial function of the yaw pressure, which we evaluate at $P_{23} = 0$ to determine the respective null parameters, F_1 and F_2.

Figures 2 and 3 plot the calibration parameters F_1 and F_2 as functions of the pitch angle for a hemispherical probe (Fig. 1D) and a conical probe (Fig. 1E). The circles (○) are data taken at the 11 different velocities ranging from 5 m/s to 30 m/s. For both probes, F_1 is nearly independent of velocity, but F_2 exhibits a small, systematic velocity dependence. The solid lines (—) are curves fitted to the data. The pitch angle (α) is fitted by a 6th degree polynomial of independent variable F_1, and F_2 is fit to 6th degree polynomial of α.

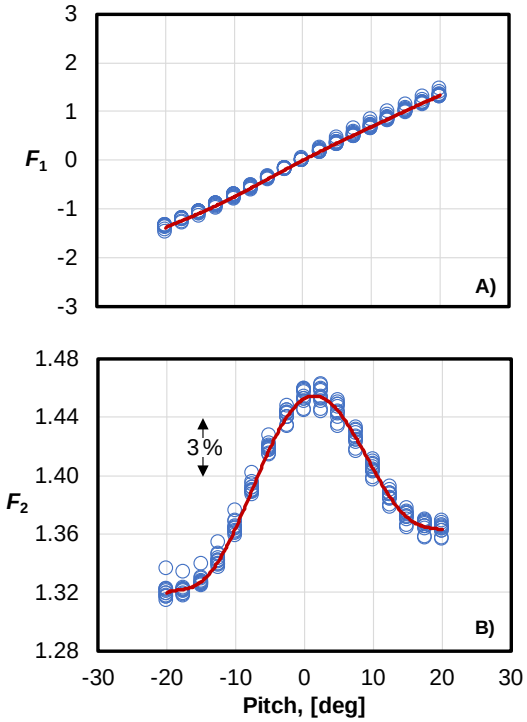


Figure 2. Hemispherical probe F_1 and F_2 calibration parameters plotted versus pitch angle. The circles (○) are data points taken at 11 different velocities and the solid line (—) is a curve fitted to the points.

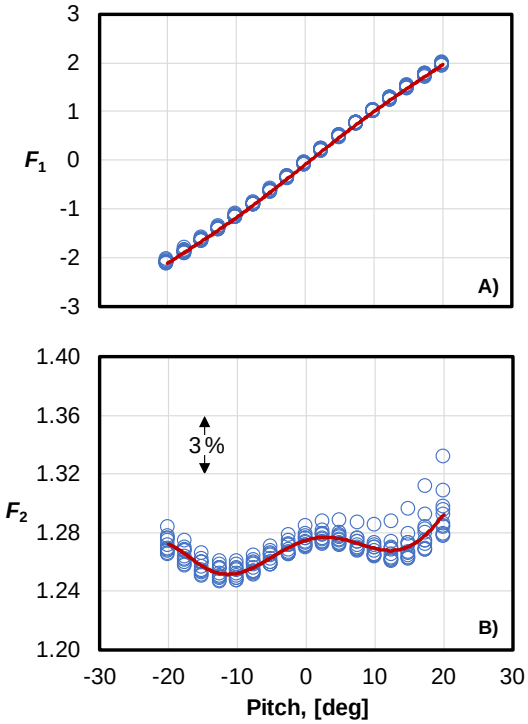


Figure 3. Conical probe F_1 and F_2 calibration parameters plotted versus pitch angle. The circles (○) are data points taken at the 11 different velocities and the solid line (—) is a curve fitted to the points.

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

As observed in Figs. 2B and 3B, the curve fit of F_2 is essentially an average of the velocity data at each pitch angle. This approximate method of accounting for the velocity dependence is consistent with the Method 2F protocol.

For flow RATAs performed using Method 2F, we determined the axial velocity at each traverse point using the following procedure. First, we nulled the probe and measured the yaw-null angle (β_{null}) with an inclinometer. Next, we determined the pitch calibration factor, $F_1 = P_{45}/P_{12}$, from the measured null pressures P_{45} and P_{12}. We use the 6th degree polynomial determined during calibration, $\alpha = \alpha(F_1)$ (here expressed in generic functional form) to determine α. Then, the calculated α is used to determine the velocity calibration factor using the fitted curve $F_2 = F_2(\alpha)$ developed during calibration. The differential pressure between ports 1 and 2 on the probe head along with the velocity calibration factor determine the dynamic pressure, $P_{dyn} = F_2^2 P_{12}$. Finally, the axial velocity at each traverse point is determined as a function of the 1) dynamic pressure, 2) yaw-null angle, and 3) pitch angle using

$$V_{axial} = \sqrt{\frac{2P_{dyn}}{\rho}} \cos\left(\beta_{null} - \beta_0\right) \cos\left(\alpha\right) \quad (1)$$

where β_0 accounts for any yaw angle offset (or misalignment) when probes are installed into the automated traverse system used to perform the flow RATA. We followed EPA Method 4 to measure the flue gas moisture [15], and we used EPA Method 3A to determine the molar mass [16]. The flue gas density (ρ) was determined *via* Method 2F using pressure, temperature, and molar mass measurements.

Non-Nulling Probe Calibrations

The non-nulling method also uses Eq. (1) to determine the axial velocity at each traverse point. The fundamental difference is that P_{dyn}, β_{null}, and α are determined by fitting 3000 or more data points acquired in NIST's wind tunnel. These data span velocities from 5 m/s to 30 m/s, pitch angles from -20° to 20°, and yaw angles from -42° to 42°. The fitted calibration curve is a fifth-degree polynomial of the four independent variables: P_{12}, P_{13}, P_{14}, and P_{15}.

For the non-nulling method, there is no need to rotate the probe. However, since scenarios could arise where rotating the probe is beneficial (*e.g.*, the predicted value of β_{null} exceeds the curve fit limits), we discuss a more general application of the non-nulling method. First the probe is rotated to a user-selected yaw angle (β). Next, we simultaneously measure the four input pressures: P_{12}, P_{13}, P_{14}, and P_{15}, and use the non-nulling calibration curve fits to calculate P_{dyn}, β'_{null}, and α. Here, β'_{null} is the calculated yaw-null angle relative to the rotated probe position at β. The absolute yaw-null angle is the sum of the probe yaw angle and the yaw-null angle determined from the non-nulling algorithm,

$$\beta_{null} = \beta + \beta'_{null} . \quad (2)$$

If the probe is oriented at a zero yaw angle ($\beta = 0°$), then the yaw-null angle determined by the non-nulling algorithm equals the yaw angle measured from the stack axis, $\beta'_{null} = \beta_{null}$. Alternatively, if one rotates the probe to the yaw-null angle, $\beta = \beta_{null}$, then β'_{null} would be zero, ideally. In this case any changes in β'_{null} would provide an indication of how the yaw-null angle fluctuates while the probe is oriented at the yaw-angle.

Test Protocol for Stack Flow Measurements

We conducted 16-point flow RATAs using multiple probe types. We used a set of 4 spherical probes (see Fig. 1C), and we also used a combination of the 2 custom probes shown in Fig. 1D and 1E. We tested each probe type at 2 loads, a high load with a nominal flow velocity of 16 m/s, and a low load of 7 m/s. The test matrix shown in Table 2 lists the probes used for each test, the flow loads, and the number of repeated runs. The diagram in Fig. 4 shows the cross-sectional view of the setup. The probe installed in each port measures the axial velocity of the nearest 4 points as illustrated in the figure. A complete traverse, herein called a *run*, includes all 16 points shown in the figure. We completed 4 runs for each probe type at the high load and 6 runs for each probe type at the low load.

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

Table 2. Test matrix for 16-point flow RATAs performed in CFPP stack.

Test No.	Probe Types	Load	Repeat Runs	Port 1	Port 2	Port 3	Port 4
1	Spherical Probes	High	4	Sphere 2	Sphere 3	Sphere 5	Sphere 6
2	Custom Probes	High	4	Hemi-sphere 1	Conical 1	Hemi-sphere 1	Conical 2
3	Custom Probes	Low	6	Hemi-sphere 1	Conical 1	Hemi-sphere 1	Conical 2
4	Spherical Probes	Low	6	Sphere 2	Sphere 3	Sphere 5	Sphere 6

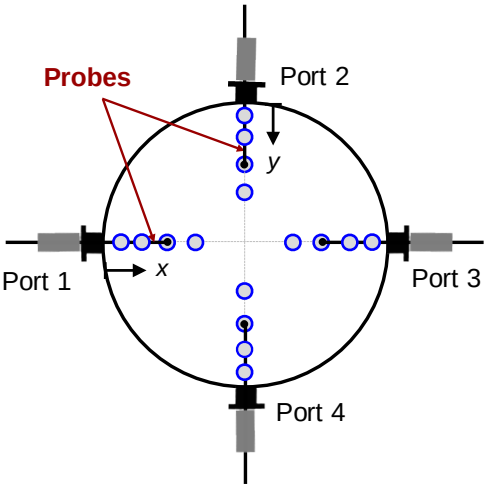


Figure 4. Cross-section of stack showing probes, port numbers and 16 traverse points located at centroids of equal stack area.

Our test protocol was conducted by an EPRI contractor who used commercially available RATA equipment called "Multiple Automated Probe System" (MAP)[1] to perform five functions: 1) move all 4 probes simultaneously to specified points; 2) periodically supply high pressure gas to purge droplets or particles plugging any of the 5 pressure ports on the probe head; 3) send a dc voltage to our data acquisition system 5 s prior to starting a purge, 4) implement the Method 2F nulling procedure including the measurement of β_{null} and β_0; and 5) provide time stamps at the start and stop of each non-nulling and Method 2F measurement intervals.

Data Acquisition System

To collect non-nulling and Method 2F data, we designed and assembled four custom data acquisition systems that were connected to a single laptop computer. Each system included inexpensive, industrial-grade differential pressure transducers, which we sampled at 10 Hz. The transducers were bidirectional with a full-scale of 1244 Pa and a time response faster than 1 kHz. We used pneumatically actuated valves to isolate the differential pressure transducers during purge events. The transducers and valves for each system were housed in a weather-proof case. Each case was placed on the floor of the RATA measurement platform just below the port where the corresponding probe was installed. Each case contained 5 pressure transducers that were connected to the 5 pressure ports on the 3-D probe using 6.35 mm inner diameter tubes approximately 13 m long. In this way, we measured the flue gas pressure (minus a near ambient reference pressure, P_{ref}, located inside the case) at all 5 pressure ports on the probe

1 Certain commercial equipment, instruments, or materials are identified in this report to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

head. The required differential pressures for the non-nulling algorithm (*i.e.,* P_{12}, P_{13}, P_{14}, P_{15}) and for Method 2F (*i.e.,* P_{23}, P_{12}, P_{45}) were calculated by subtracting the appropriate pressure measurements. For example, the yaw pressure was determined by subtracting the measured pressures on port 2 from port 3, $P_{23} = (P_2 - P_{ref}) - (P_3 - P_{ref})$.

Procedure for Automated Traverses

Each of the 4 tests listed in Table 2 began by starting the data acquisition unit. Pressure data were collected throughout the test except during purge events, which occurred approximately once every minute. During purge events, valves isolated the transducers from the purge pressure while simultaneously re-zeroing the transducers to the common reference pressure.

The same measurement protocol was followed at each traverse point. The MAP system simultaneously moved the 4 probes to the specified traverse point and rotated each probe to a zero yaw angle. After a 3 s stabilization period, the axial velocity ($V_{NN@0yaw}$) was measured for 10 s using the non-nulling algorithm. Next, the MAP system nulled each probe and recorded its β_{null}. After another 3 s stabilization period we measured the axial velocity for 10 s via Method 2F (V_{M2F}) and the non-nulling method ($V_{NN@null}$). Thus, we measured 3 velocities at each traverse point: 1) non-nulling with the probe at zero yaw; $V_{NN@0yaw}$, 2) Method 2F at the yaw-null angle; V_{M2F}, and 3) a second non-nulling measurement coincident with V_{M2F} where the probe is oriented at the yaw-null angle; $V_{NN@null}$. The second non-nulling measurement provided insight regarding the steadiness of the yaw-null angle, and could be directly compared to V_{M2F} since both measurements were made simultaneously.

Data Processing

The 3 axial velocities (*i.e.*, $V_{NN@0yaw}$, V_{M2F}, and $V_{NN@null}$) measured at each traverse point are all calculated using Eq. (1). However, the algorithms for determining P_{dyn}, β_{null}, and α differ for the non-nulling method and Method 2F. The calculations for both methods are outlined above in the section entitled *Probe Calibrations*. In this section, we emphasize the different approach in averaging the data in each 10 s collection interval.

Method 2F determines the average axial velocity and pitch angle from *pressure averages.* Specifically, we calculated $P_{12,avg}$ and $P_{45,avg}$, which are arithmetic averages of P_{12} and P_{45} sampled at 10 Hz for 10 s.

In contrast, our implementation of the non-nulling method determines the average dynamic pressure (P_{dyn}), yaw-null angle (β_{null}), and pitch angle (α) from *time averages*. These quantities are calculated every 0.1 s when P_{12}, P_{13}, P_{14}, and P_{15} are updated. At the end of the 10 s collection interval, we calculate the arithmetic average of the 100 values of P_{dyn}, β_{null}, and α. As expected for the steady flows in NIST's wind tunnel, the values of P_{dyn}, β_{null}, and α computed from the pressure averages and the time averages were indistinguishable. If transients are present in the stack flow, a time average may be more accurate than a pressure average. In the CFPP stack, we compared the axial velocities V_{axial} determined from pressure averages and time averages in a few cases. For most of the comparisons, the values of V_{axial} agreed to better than 1 %; in a few cases V_{axial} differed by 10 % or more.

One disadvantage of time-averaging is that the noisy pressure signals occasionally yielded values of P_{dyn}, β_{null}, or α that exceed the limits of the non-nulling calibration curve. This problem was unexpected; we circumvented it by excluding the anomalous values from the averages. Fortunately, there were only a few cases where the calculated average did not include at least 80 % of the data. In future tests we will expand the range of the non-nulling calibration curve and we will retake data points that do not include at

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

least 80 % (or some user-specified percentage) of data in the time averages.

Unfortunately, the data acquisition was not set up to process data during the CFPP stack measurements. Therefore, we processed the data after the field tests were completed. We used the time stamps provided by the MAP system to identify the non-nulling and Method 2F pressure data. For the low loads, approximately 20 % of the data could not be found at the indicated time stamps. At the high load less than 5 % of the data was unaccounted for.

Results

Table 1 summarizes the average flow results. It provides solid evidence that the non-nulling method has the potential to make efficient, accurate stack flow measurements. Because we already discussed the averaged flow data, we now compare the profiles of velocity, yaw-null angle, and pitch angle determined by Method 2F to those determined by the non-nulling method. In addition, we describe how we used the noisy pressure signals to troubleshoot plugging problems.

The flow RATAs were performed along 2 orthogonal axes. We denote the axis extending from port 1 to port 3 in Fig. 4 as the "*x*-axis". The *y*-axis extended from port 2 to port 4. Each axis included 8 traverse points. The traverse points are located at the centroids of equal area, so that flow velocity of each run is calculated by averaging the axial velocities measured at 16 traverse points [1]. The axial velocity, yaw-null angle, and pitch angle are plotted on the *x*/*D* and *y*/*D* axes, respectively, where *D* is the diameter of the stack.

Axial Velocity Profiles

Figures 5A and 5B show that while the load remained constant, the flow profile had large variations (greater than 10 %) at particular locations. Figure 5 is a plot of the normalized axial velocity (V_{RATA}/V_{CEMS}) measured using the spherical probes at high load as functions of *x*/*D* and *y*/*D*, respectively. The open circles (○) connected by dashed lines are Method 2F data from each of the 4 runs. The spacings between the dashed lines indicate profile variations. Despite these variations, the flow velocity of each Method 2F run is stable as shown in Table 3. The standard deviation expressed as a percent was only 2.1 %.

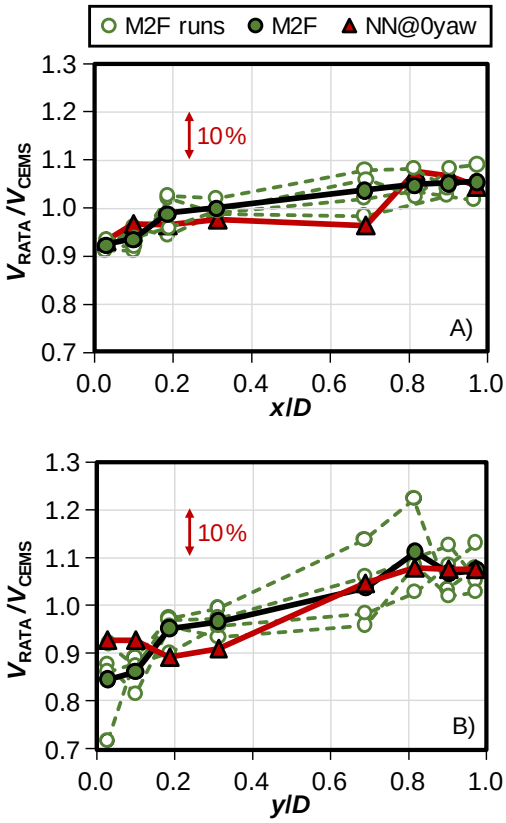


Figure 5. Flow RATA for spherical probes at high load: Plots of normalized axial velocity versus A) *x*/*D,* and B) *y*/D.

We observed similar profile variations (not plotted) in the 4 non-nulling runs even though the standard deviation of the average velocity was only 0.4 %.

The localized variations in the flow field indicated in Figs. 5A and 5B might be due to vortices. We are confident that they are not artefacts of the measurements (*e.g.*, caused by plugging or filtering the data) because the average flow velocity for each run is stable.

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

The solid circles (●) and solid triangles (▲) in Fig. 5 are the averages of the Method 2F runs and the non-nulling runs, respectively. In Figs. 5A and 5B the solid lines connecting the averaged points are close to each other. This displays the good agreement of the Method 2F velocity profiles with the non-nulling velocity profiles. Table 3 shows that the difference in the averaged flow velocity is only -0.1 %. We emphasize that the normalized velocity profiles measured at both high and low loads were similar to the profiles observed in Figs. 5A and 5B.

Table 3. Normalized flow velocities determined by Method 2F and by the non-nulling method for the 4 repeated runs measured with spherical probe at high load (16 m/s).

Run No.	$\frac{V_{M2F}}{V_{CEMS}}$	$\frac{V_{NN@0yaw}}{V_{CEMS}}$	% Diff[c]
1	1.008	0.999	0.9 %
2	1.009	0.993	1.6 %
3	0.965	0.991	-2.6 %
4	0.988	0.992	-0.4 %
Avg[a]	**0.993**	**0.994**	**-0.1 %[c]**
%Std Dev[b]	**2.1 %**	**0.4 %**	

a) Avg is the average of the 4 runs

b) %Std Dev is 100 times the standard deviation of the 4 runs dividing by the average

c) %Diff is calculated by $100(V_{M2F}/V_{NN@0yaw}-1)$

Yaw Angle Profiles

Figure 6 shows the average yaw-null profiles for the spherical probe at high load. The Method 2F (●) and non-nulling (▲) yaw-null angles show the same trend and are in good agreement in Figs. 6A and 6B. Both methods show the magnitudes of yaw-null angles are largest near the wall with a value of nearly 30°. The magnitude yaw-null angle decreases monotonically as one moves away from the wall toward the center of the stack. The differences between Method 2F and the non-nulling method are smallest near the center of the stack and increase to maximum of approximately 7° near the wall in the worst case.

We found that the yaw-null profiles were nearly identical at low load. We obtained the same trends shown in Figs. 6A and 6B independent of probe type (*i.e.*, spherical or custom) and method (*i.e.*, non-nulling or Method 2F).

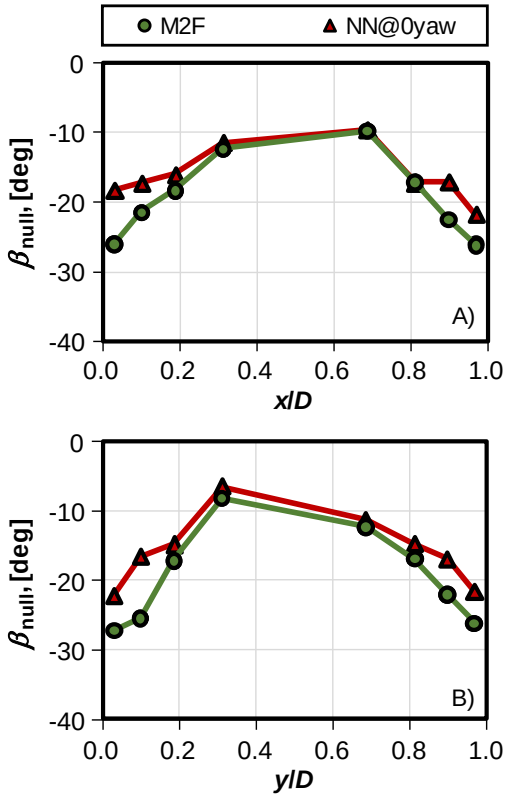


Figure 6. Yaw-null profiles determined using Method 2F (●) and non-nulling with $\beta = 0°$ (▲) along A) port 1 to port 3, and B) port 2 to port 4.

Figure 7 plots the yaw-null angle during a typical 10 s collection time with the probe oriented at $\beta = \beta_{null}$. Because the probe was nulled, the non-nulling algorithm measures β'_{null} defined by Eq. (2). In a steady flow with low turbulence β'_{null} would have a constant value close to 0° during the 10 s collection. In contrast, we observed (Fig. 7) the sine-like oscillations with an amplitude of nearly 30°

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

and a period of approximately 4 s. Surprisingly, the integrated average of β'_{null} is -1.5°, which is close to zero. This time-dependence of β'_{null} is evidence that the flow field in the CFPP stack had large transients. (Figs. 5A and 5B are additional evidence for large transients.) We note that better averages could be obtained by averaging over more cycles (*i.e.*, longer collection times) or by averaging over the 4 s period. However, since the focus of this work is to demonstrate the feasibility of the non-nulling method, we did not implement this strategy.

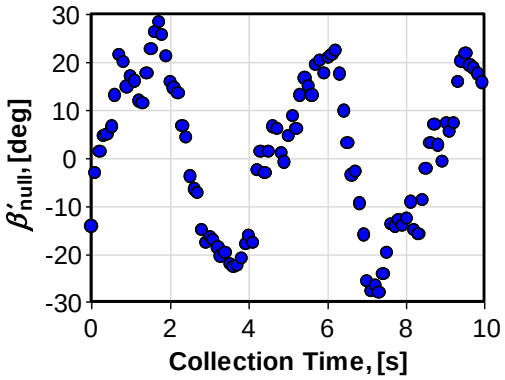


Figure 7. Sine-like oscillations of Yaw-null angle during 10 s Method 2F data collection. (Spherical probe is oriented at the yaw-null angle, and β'_{null} is determined every 0.1 s using the non-nulling algorithm.)

Pitch Angle Profiles

Figures 8A and 8B show profiles of the pitch angle determined by Method 2F (●) and by the non-nulling algorithm with $\beta = 0°$ (▲). These results correspond to Test #1 specified in Table 2. The pitch angles determined by Method 2F and by the non-nulling algorithm agree with each other and have similar, asymmetric dependences on *x*/*D* and *y*/*D*. We found the same characteristic profiles independent of flow load, probe type, and method. Although we hoped to perform the test in a stack with high pitch, the largest pitch angle was only about 5°.

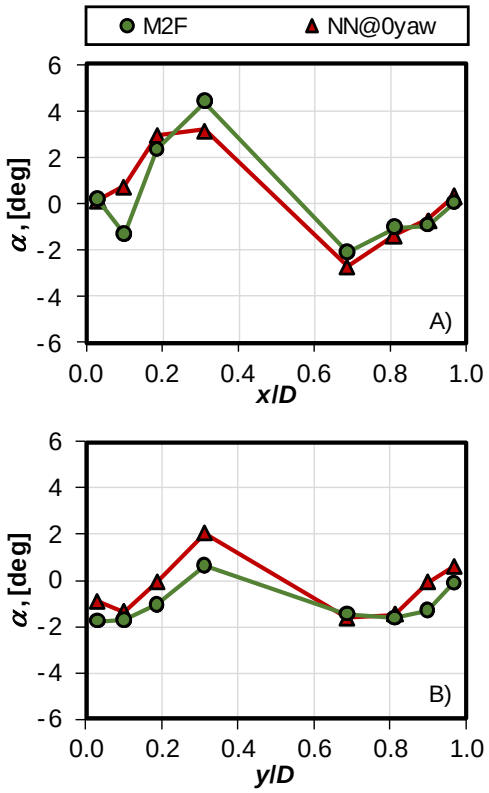


Figure 8. Pitch angle profile for Test #1 in Table 2 where A) *x*/*D* is the dimensionless distance from port 1 to port 3, and B) *y*/*D* is the dimensionless distance from port 2 to port 4.

Troubleshooting Plugging Problems

To mitigate plugging we purged the probe pressure ports every 60 s. Nevertheless, we still had problems with plugging. Plugging issues were most severe for spherical probe 2 during Test #4 in Table 2. The 4 traverse points in port 1 seemed to be the most impacted by plugging problems.

One way to detect plugging is to evaluate the consistency of repeated axial velocity measurements made at the same traverse point. If significant deviations are found at the same traverse points from run to run, then plugging could be the culprit. However, we could not be sure if deviations resulted from plugged pressure ports or from flow transients like those observed in Figs. 5A and 5B. In this

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

study we used a simple statistical approach to find outliers in the data caused by plugging.

The pressure signals ($P_{n,ref}$; n = 1 to 5) for the five pressure ports on the probe head were noisy. That is, pressures fluctuations during non-nulling and during Method 2F were usually larger than the mean of the pressure signal. We hypothesized that the noise would significantly decrease if a pressure port on the probe head was plugged. For each 10 s collection time, we computed the standard deviation of the pressure signal from each pressure port on the probe head. If the standard deviation was below the typical noise level by a statistically defined threshold, we assumed that the port was plugged.

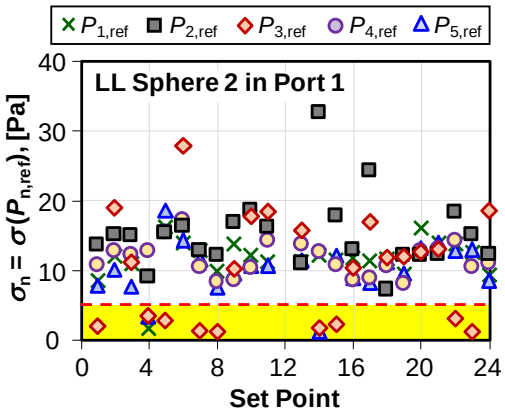


Figure 9. Standard deviation of pressure signals (σ_n) at n = 1 to 5 pressure ports on the spherical probe head. Values of σ_n below the dashed line (– –) indicate that port n was plugged.

Figure 9 illustrates how we applied the statistical approach to detect plugged pressure ports. This example focuses on the non-nulling measurements made at low load using spherical probe 2. The 24 set points on the *x*-axis correspond to the 4 traverse points for port 1 multiplied by the 6 repeated runs (see Test #4, Table 2). The *y*-axis is the standard deviation of the pressure signals $\sigma_n = \sigma(P_{n,ref})$ measured at the n = 1 to 5 pressure ports on the probe head. We considered a pressure port plugged if the standard deviation was below the statistical limit indicated by the dashed line (– –). For simplicity Fig. 9 only shows a single limit; however, in practice we used separate limits for each of the 5 pressure signals. The statistical limit for the nth probe was

$$limit_n = \langle \sigma_n \rangle - k\,\sigma(\sigma_n) \qquad (3)$$

where $\langle \sigma_n \rangle$ is the average of the 24 values of σ_n; $\sigma(\sigma_n)$ is the standard deviation of the 24 values of σ_n; and k is the coverage factor which we set equal to 1.5. (The computed normalized velocities had only a weak sensitivity to the value of k.)

Figure 10A compares two normalized velocity profiles, one affected by plugging, and the other calculated excluding the subset of data affected by plugging. The figure corresponds to traverses performed at low load using the spherical probes. The velocity (V_{RATA}) was determined using the non-nulling algorithm with the probe oriented at zero yaw angle. Each open triangle (△) is the average of 6 repeated runs. The dashed line connecting the triangles shows the normalized axial velocity profile of the 8 traverse points between port 1 and port 3 (*i.e.*, the x-axis). The first 4 points along *x*/D are traversed by the spherical probe 2 installed in port 1. The statistical approach illustrated in Fig. 9 suggested that several of these points were affected by plugging. If we omit these points when calculating the average axial velocity at each traverse point, we obtain the solid triangles (▲). The solid line connecting the solid triangles shows the normalized velocity profile corrected to account for plugging.

If the normalized velocity profile (▲) in Fig. 10A is correct, we expect to find the same profile at low load independent probe type (*i.e.*, spherical or custom) and independent of the method (*i.e.*, non-nulling or Method 2F). Moreover, for these high Reynolds number flows (3×10^6 to 6.5×10^6) we expect that the high load normalized velocity profile will have essentially the same shape as the low load. Figure 10B

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

shows that all normalized profiles are in good agreement with the corrected profile. The good agreement of these profiles is 1) strong evidence that we successfully identified and removed data affected by plugging, and 2) that the non-nulling method performed well independent of probe type and flow load.

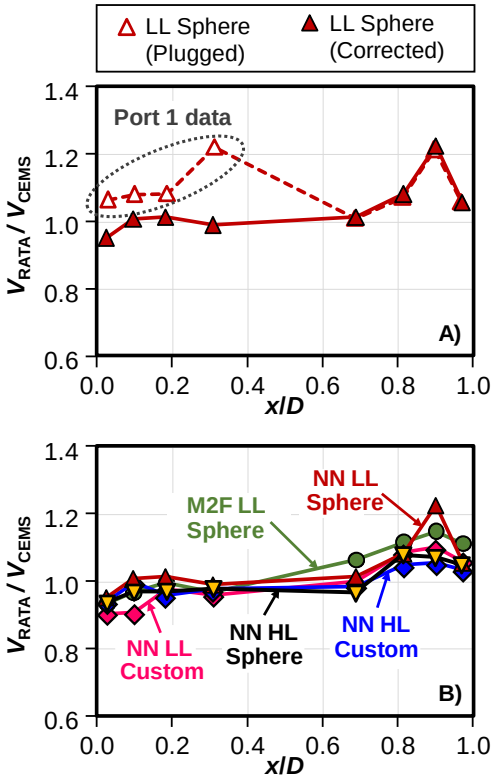


Figure 10. Normalized axial velocity profiles plotted against the dimensionless distance from port 1 to port 3: A) △ Low Load Spherical profile with plugged probe ports; ▲ same profile recalculated with plugged data removed.
B) Five profiles not significantly affected by plugging:
1) ◆ NN LL Custom non-nulling low load,
2) ◆ NN HL Custom non-nulling high;
3) ▼ NN HL Sphere non-nulling high load,
4) ● M2F LL Sphere, Method 2F low load
5) ▲ NN LL Sphere, non-nulling, low load

Conclusions

We demonstrated that the non-nulling method can accurately measure complex flows in CFPP stacks. We conducted 16-point flow RATAs 3.8 stack diameters downstream of the 90° elbow at the stack inlet, and we measured yaw-null angles approaching –30° near the stack wall. We found excellent agreement between the non-nulling method and Method 2F using spherical probes. The results from Table 1 show agreement of -0.1 % at a high load of 16 m/s and 0.0 % at a low load of 7 m/s. We found similar levels of agreement between Method 2F and the non-nulling method when we conducted flow RATAs in NIST Scale-Model Smokestack Simulator (SMSS) [7, 8]. The non-nulling method gives the same flow results but is more time and cost efficient than Method 2F.

The SMSS facility uses air as a surrogate for flue gas and has a 1.2 m diameter test section. The facility can generate complex flows that have yaw-null angles of almost 40° at the wall. The excellent non-nulling flow results found in the SMSS are analogous to those found in this study of a CFPP stack. Thus, the SMSS facility is a satisfactory research facility for characterizing probes used for flow RATAs, ultrasonic CEMS, and other flow monitors for use in CFPP stacks.

We developed custom hemispherical and conical probes and compared their performance in a CFPP stack with the EPA-sanctioned spherical probe using the non-nulling method. The non-nulling flow velocities at high and low loads were consistent for all probe types. After normalizing the measured axial velocities by the CEMS velocity, we found essentially the same characteristic profiles at low and high loads across both orthogonal chords. The normalized Method 2F axial velocities also exhibited the same profiles across the chords.

The non-nulling method measured consistent pitch and yaw-null angles using all the probe types at both high and low loads. Therefore, in future flow RATA testing, a hybrid non-nulling method can be implemented. That is, if while performing a flow RATA using the non-nulling method one has reason to question the axial velocity measurement, the RATA tester can

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

rotate the probe to the calculated yaw-null angle and take a Method 2F measurement.

The non-nulling method requires bidirectional, fast response differential pressure transducers. We used industrial grade differential transducers for our stack measurements. We measured the pressure (minus a common reference pressure) at each of the 5 ports on the 3-D probe. Pressure measurements were sampled at 10 Hz. They revealed periodic pressure fluctuations with periods ranging between 3 s and 5 s. These transients could not be observed or adequately accounted for (*e.g.*, averaging over the periods) using Method 2F. In contrast, the non-nulling data processing could easily be modified to perform averages over the period.

We used a commercially available automated traverse system 1) to reduce the RATA times and 2) to improve the accuracy of nulling the probes. We emphasize that the benefits of automated traverses are less important for the non-nulling method than for nulling methods because the non-nulling method does not rotate the probe rotation and eliminates errors from imperfect nulling.

Despite purging every 60 seconds, our spherical probes were plagued by plugging that was most severe at low load. We did not experience the same difficulties with the two custom probes; however, we are not sure if this is due to their designs or to good fortune. Additional field tests are needed to better understand plugging.

For accurate flow measurements, it was necessary to distinguish fluctuations of the axial velocity from plugging of one or more pressure ports. We made this distinction by detecting the reduction in the pressure noise that occurs when a pressure port is plugged. Without plugging, the fluctuations of the pressure signals were often larger than their mean values. For each pressure signal, during each 10 s data collection period, we used the standard deviation of the pressure from its mean as a measure of its noise. In this study, we were not prepared to process the noise data in real time. After all the measurements were completed, we used a statistical criterion to discard data corrupted by plugging. In the future, we will process the noise data as they are acquired during a RATA. If the noise indicates plugging the probe can be purged and the data retaken. Thus, the noise measurements will be used as a diagnostic to guide the data acquisition and not to discard data.

In this study we performed a 3000-point calibration on each probe used for the non-nulling measurements. Such an extensive calibration is not practical for routine flow RATAs. Research efforts are underway to determine if a baseline non-nulling calibration can be applied to all probes of the same type. If so, a simple calibration will be done to correct for slight manufacturing differences.

Acknowledgments
We thank John Wright of NIST for his continuing, enthusiastic support of this work. This research was partially funded by NIST's Greenhouse Gas and Climate Science Measurements Program. A Cooperative Research and Development Agreement (CRADA) between NIST and EPRI facilitated our access to a RATA at a CFPP.

[1] Environmental Protection Agency, ***40 CFR Part 60 Application Method 1 Sample and Velocity Traverses for Stationary Sources***, Washington D.C., 1996.

[2] Norfleet, S. K., Muzio L. J., and Martz T. D., ***An Examination of Bias in Method 2 Measurements Under Controlled Non-Axial Flow Conditions***, Report by RMB Consulting and Research, Inc., *May 22, 1998.*

[3] Environmental Protection Agency, ***40 CFR Part 60 Application Method 2 Velocity and S-type Pitot***, Washington D.C., 1996.

[4] Environmental Protection Agency, ***40 CFR Part 60 Application Method 2G Stack***

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

Gas-Velocity/Volumetric Flow Rate-2D probe, Washington D.C., 1996.

[5] Environmental Protection Agency, 40 CFR Part 60 ***Application Method 2F Stack Gas-Velocity/Volumetric Flow Rate-3D probe***, Washington D.C., 1996.

[6] Shinder, I. I., Johnson, A. N., Moldover, M. R., Filla, B. J., and Khromchenko, V. B., ***Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing,*** 10th International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico, MX, April 2018.

[7] Johnson, A. N., Boyd, J. T., Harman, E., Khalil, M. Ricker, J. R., Crowley, C. J., Wright, J. D., Bryant, R. A., and Shinder, I. I., **Design and Capabilities of NIST's Scale-Model Smokestack Simulator (SMSS),** 9th International Symposium on Fluid Flow Measurement (ISFFM), Arlington, Virginia, April 2015.

[8] Bryant, R. A.; Johnson, A. N.; Wright, J. D.; Wong, T. M.; Whetstone, J. R.; Moldover, M. R.; Shinder, I. I.; Swiggard, S.; Gunning, C.; Elam, D.; Martz, T.; Harman, E.; Nuckols, D.; Zhang, L.; Kang, W.; and Vigil, S.; ***Improving Measurement for Smokestack Emissions - Workshop Summary***, Special Publication (NIST SP) 1228, 9/21/2018.

[9] Johnson, A. N., Shinder, I. I., Moldover, M. R., Boyd, J. T., and Filla, B. J., ***Progress Towards Accurate Monitoring of Flue Gas Emissions*****,** 10th International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, MX, March 2018.

[10] Yeh, T. T., Hall, J. M., *Air Speed Calibration Service*, NIST Special Publication 250-79, National Institute of Standards and Technology, Gaithersburg, Maryland, 2006.

[11] Yeh, T. T., Hall, J. M., ***An Uncertainty Analysis of the NIST Airspeed Standards, ASME Paper FEDSM2007-37560***, ASME/JSME 5th Joint Fluids Engineering Conference, pp. 135-142, San Diego, California, USA, 2007.

[12] Yeh, T. T. and Hall, J. M. ***Uncertainty of NIST Airspeed Calibrations***, *Technical document of Fluid Flow Group*, NIST, Gaithersburg, Maryland, USA, 2008.

[13] Shinder, I. I., Crowley, C. J., Filla, B. J., Moldover, M. R., ***Improvements to NIST's Air Speed Calibration Service***, 16th International Flow Measurement Conference, Flomeko 2013, Paris, France September 24-26, 2013.

[14] Shinder, I. I., Khromchenko, V. B., and Moldover, M. R., ***NIST's New 3D Airspeed Calibration Rig, Addresses Turbulent Flow Measurement Challenges,*** 9th International Symposium on Fluid Flow Measurement (ISFFM), Washington, DC, U. S., April 2015.

[15] Environmental Protection Agency**, *40 CFR Part 60 Application Method 4 Determination of Moisture Content in Stack Gases***, Washington D.C., 1996.

[16] Environmental Protection Agency**, *40 CFR Part 60 Application Method 3A Determination of Oxygen and Carbon Dioxide Concentrations in Emissions From Stationary Sources***, Washington D.C., 1996.

Johnson, Aaron; Shinder, Iosif; Filla, Bernard; Boyd, Joey; Bryant, Rodney; Moldover, Michael.
"Non-Nulling Measurements of Flue Gas Flows in a Coal-Fired Power Plant Stack."
Paper presented at Flomeko 18th International Flow Measurement Conference, Lisbon, Portugal. June 26, 2019 - June 28, 2019.

Security Awareness in Action: A Case Study

Julie M. Haney, *National Institute of Standards and Technology*
Wayne G. Lutters, *University of Maryland*

Abstract

A critical role and force-multiplier in security adoption is the cybersecurity advocate. Cybersecurity advocates are security professionals who attempt to remedy implementation failures by actively promoting and facilitating the adoption of security best practices and technologies as an integral component of their jobs. These advocates not only have technical skills, but also must possess the ability to educate, persuade, and serve as organizational change agents for cybersecurity adoption.

A prior interview study of a diverse set of cybersecurity advocates advanced the definition of the cybersecurity advocate role [1]. However, the study was limited in that it was a one-sided self-report view through the lens of advocates. Additional empirical evidence of cybersecurity advocates working "in the wild" (within an actual work context) is needed to validate the findings. To obtain a more comprehensive look at cybersecurity advocacy in practice, we are conducting an in-depth case study of a security awareness team at a mid-sized U.S. federal government agency.

Security awareness professionals are a type of cybersecurity advocate who are responsible for developing and executing security awareness programs within their own organizations. Prior insight into the day-to-day work and challenges of these professionals reveal that the majority of respondents perform security awareness duties on a part-time basis with little budget, with many lacking sufficient background and soft skills (e.g., communications, relationship-building, and marketing) that are needed to effectively engage the workforce and key departmental stakeholders such as the finance and operations departments [2,3].

The in-progress case study will allow for examination of a security awareness team from several perspectives via a multi-faceted approach involving: 1) interviews of security awareness team members, managers in the team's chain-of-command, and agency employees who receive the security awareness information, 2) field observations of four in-person security awareness events, and 3) review and analysis of security awareness materials distributed to the agency's workforce.

Preliminary qualitative analysis reveals a passionate, creative team willing to try new approaches to attain their goal of making security awareness entertaining and informative rather than a mandatory burden. We will discuss the techniques and approaches of the team and how their efforts are viewed by others at their agency. Examples of the team's approaches include: ensuring security awareness information is timely and topical to the season, world events, or current organizational concerns; providing employees with security awareness information that is not only relevant to their jobs, but also to their personal life; introducing humor and gaming when appropriate; and soliciting feedback from others in the organization about what topics to address in future events. We will also discuss challenges security awareness professionals face, including lack of resources and employee attitudes and time constraints.

As a complement to the prior cybersecurity advocate interview study, the case study will provide better insight into real-world security advocacy techniques and which are most effective. These practices may then inform the design of security interfaces and training. In addition, the project allows for a deeper examination of professional characteristics of advocates, and how those are viewed by advocates' target populations. The identification of these characteristics can aid in the creation of professional development resources to aid people in becoming successful advocates.

References

[1] J. M. Haney and W. G. Lutters. "It's Scary…It's Confusing…It's Dull": How cybersecurity advocates overcome negative perceptions of security. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 411-425, 2018.

[2] SANS. 2018 Security Awareness Report. https://www.sans.org/sites/default/files/2018-05/2018%20SANS%20Security%20Awareness%20Report.pdf, 2018.

[3] B. Woelk. The successful security awareness professional: Foundational skills and continuing education strategies. https://library.educause.edu/~/ media/files/library/2016/8/erb1608.pdf, 2015.

Workshop on Security Information Workers, USENIX Symposium on Usable Privacy and Security (SOUPS) 2019. August 11, 2019, Santa Clara, CA, USA.

.

Haney, Julie; Lutters, Wayne.
"Security Awareness in Action: A Case Study."
Paper presented at WSIW 2019 - 5th Workshop on Security Information Workers (held at the 15th Symposium on Usable Privacy and Security), Santa Clara, CA, United States. August 11, 2019 - August 11, 2019.

Perceptions of Smart Home Privacy and Security Responsibility, Concerns, and Mitigations

Julie Haney
National Institute of Standards and Technology
julie.haney@nist.gov

Susanne Furman
National Institute of Standards and Technology
susanne.furman@nist.gov

Yasemin Acar
Leibniz University Hannover
acar@sec.uni-hannover.de

Mary Theofanos
National Institute of Standards and Technology
mary.theofanos@nist.gov

Poster presented at the 15th Symposium on Usable Privacy and Security (SOUPS 2019)

Abstract
Smart home devices are increasingly being used by non-technical users who have little understanding of the privacy and security implications of the technology. To better understand perceptions of smart home privacy and security, we are conducting an interview study of individuals living in smart homes. Preliminary analysis reveals potential relationships between perceptions of responsibility and privacy and security concerns and mitigation actions. Results can inform future efforts to educate users about their responsibility, advance the protection of user data, and protect the devices from unintended access.

Author Keywords
Internet of things; smart home; usable security; usable privacy

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
The Internet of Things (IoT) market is exploding, with the number of IoT devices expected to grow from 26 billion in 2019 to 75 billion in 2025 [5]. With this growth, IoT technology is becoming more pervasive in the home environment, with 34% of broadband households forecasted to have

Haney, Julie; Furman, Susanne; Theofanos, Mary; Acar Fahl, Yasemin.
"Perceptions of Smart Home Privacy and Security Responsibility, Concerns, and Mitigations."
Paper presented at WSIW 2019 - 5th Workshop on Security Information Workers (held at the 15th Symposium on Usable Privacy and Security), Santa Clara, CA, United States. August 11, 2019 - August 13, 2019.

Privacy Concerns: *"If somebody has access to this cloud information and they're actually able to associate when you're home and when you're not home based on the sensors and other things you have in your house, they could potentially target you."* (P11_A)

Security Concerns: *"what if someone hacks into our phones and. . . with the Nest. . . they try to reconfigure it on their own or especially the alarm system."* (P12_U)

Lack of Concern: *"I feel like you've got people who are pretty talented with computers and can get this stuff. But again, I'm of the mindset, have at it. We don't do anything cool in my house, anyways."* (P8_A)

Tradeoffs: *"I know that it's collecting personal data. . . and I know there's the potential of a security leak, but yet, I like having the convenience of having those things."* (P1_A)

smart home systems by 2025 [1]. While early adopters of smart home technology have typically been more technically savvy, IoT smart home devices are increasingly being used by non-technical users who have little understanding of the technology or awareness of the implications of use. In particular, the impact of and interplay of factors such as usability, security, privacy, and trust have not been adequately explored within one comprehensive study. We address this gap with an in-progress qualitative interview study of individuals living in smart homes. Preliminary analysis focused on privacy and security reveals potential relationships between perceptions of responsibility, concerns, and mitigation actions. Results can inform future efforts to educate users about their privacy and security responsibility, advance the protection of user data via usable interfaces, and protect the devices from unintended access.

Related Work

Prior work has examined perceptions of smart home privacy and security. Parks Associates [2] and Worthy et al. [7] found that a lack of trust in vendors to properly safeguard personal data is a major obstacle to adoption of smart home technology. From a broader IoT perspective, Williams et al. [6] found that IoT is viewed as less privacy-respecting than non-IoT devices such as desktops, laptops, and tablets. However, users' privacy concerns were not always translated into privacy-protecting actions. Interviews of people living in smart homes by Zeng et al. [8] and a PwC industry survey [4] revealed that, although users may be aware of security and privacy issues, these were often overlooked when a product proved otherwise valuable.

Methods

We conducted 15 semi-structured interviews, lasting on average 50 minutes, as part of an in-progress study to understand end users' perceptions of and experiences with smart home devices. The study was approved by the NIST Human Subjects Protection Office. Prospective participants first completed an online screening survey about their smart home devices, their role with the devices (e.g., purchaser, administrator, user), and professional backgrounds. Participants were selected for interviews if they had multiple smart home devices for which they were an active user. Of the 15 participants, 12 had installed and administered the devices (indicated with an A after their participant ID) and three were non-administrative users of the devices (indicated with a U).

Interview questions addressed several areas: understanding of smart home terminology; purchase and general use; installation and troubleshooting; privacy; security; and safety. Interviews were audio recorded and transcribed. We then performed iterative coding and qualitative analysis on the data to identify core concepts [3]. In this poster, we report on a subset of our preliminary findings specific to privacy and security concerns, mitigations, and responsibility.

Preliminary Findings

Preliminary analysis reveals possible relationships between privacy and security concerns, enactment of mitigations to alleviate those concerns, and perceptions of responsibility for the privacy and security of smart home devices. Example quotes for each concept are provided in the side bars.

Concerns

Participants were asked if they had any hesitations or concerns about their smart home devices during which time many participants, unprompted, discussed privacy or security. They were later asked explicitly about their privacy and security concerns. All participants acknowledged privacy concerns (12 unprompted). Concerns were either personally held or those they had heard others express, with 11

Haney, Julie; Furman, Susanne; Theofanos, Mary; Acar Fahl, Yasemin.
"Perceptions of Smart Home Privacy and Security Responsibility, Concerns, and Mitigations."
Paper presented at WSIW 2019 - 5th Workshop on Security Information Workers (held at the 15th Symposium on Usable Privacy and Security),
Santa Clara, CA, United States. August 11, 2019 - August 13, 2019.

Privacy Responsibility

Personal responsibility: *"if you want your information protected, don't bring a camera into your house. If you want your information not to be put on the cloud, don't give the opportunity. So, I think that you are ultimately accountable. It's your information."* (P8_A)

Manufacturer responsibility: *"[Those responsible are] really the people who write the software that interact with the devices, people that write the firmware for the devices, and then the software that essentially runs in the cloud services that aggregates all of this data, and then performs functions. So you're really at their mercy."* (P11_A)

Government responsibility: *"voluntary consensus on privacy issues is almost impossible to get from the commercial sector. . . I think they need privacy guidelines at least from the government in order to adhere to them."* (P13_A)

being personally concerned. Fourteen voiced security concerns (11 unprompted, 10 personally concerned).

Despite having concerns, participants were more than willing to bring devices into their homes. For example, one participant commented, *"it's not gonna stop me from living my life. . . But we do take it into consideration the privacy aspects of things, but it's not to any extreme" (P6_U).*

Mitigations

Concern often, but did not always, translate into action. During the privacy and security portions of the interviews, participants were asked if they performed any mitigations to alleviate their concerns. Of the 11 who were personally concerned about privacy, nine discussed implementing mitigations. The most commonly mentioned privacy mitigations were: configuring privacy-related options (e.g., not sending usage statistics, disabling ordering) (6 participants); covering/repositioning cameras (3); and not putting listening devices in rooms where sensitive conversations could occur (3). Not surprisingly, among those four who were not concerned about privacy, only one implemented a mitigation.

Eight of the 10 participants who were personally concerned about security mentioned mitigations. The most frequently mentioned security mitigations included: password management (e.g., strong passwords and changing passwords on apps) (7); home network security (e.g., secure WiFi, network segmentation) (6), configuring security options on the devices (4), choosing devices with strong security features (3), and physical security of devices (2).

Security mitigations in particular demonstrated lack of understanding of mitigation effectiveness as well as confusion about the relationship between smart home devices and other activities such as social media, web browsing, and email. For example, when asked what smart home device privacy mitigations he takes, P4_A mentioned that he does not go on Facebook and tries to clean up old emails.

Household members also influence mitigations as was the case for four participants not personally concerned about privacy or security. One participant said, *"My husband is more security minded. . . The Alexa device has a video camera that you can use, but he's taped it over" (P1_A).*

Responsibility for Privacy

Participants were asked who they thought was responsible for protecting the privacy of information collected by their smart home devices. Responses included three different entities: themselves, device manufacturers, and the government, with only one participant saying they did not know. Responsibility was often viewed as being shared.

Eight placed partial responsibility on themselves. Two of those eight put sole responsibility on themselves. One such participant did not trust device vendors since *"the manufacturers' desires are counter to the consumer" (P16_A).*

Eleven believed manufacturers share some responsibility, with four of those claiming manufacturers have sole responsibility. For example, one participant remarked *"They need to do everything [since they are] taking so much money for all that" (P9_A).* However, even while putting some responsibility on manufacturers, participants do not completely trust them. When asked if he ever reads any of the privacy agreements, P10_A said, *"I don't have much trust in what companies say they collect and don't collect. I think they collect what they can and use it" (P10_A).*

Four participants felt that the government had some responsibility to regulate smart home device privacy along with manufacturers and/or themselves. One mentioned the European Union's General Data Protection Regulation as a

Haney, Julie; Furman, Susanne; Theofanos, Mary; Acar Fahl, Yasemin.
"Perceptions of Smart Home Privacy and Security Responsibility, Concerns, and Mitigations."
Paper presented at WSIW 2019 - 5th Workshop on Security Information Workers (held at the 15th Symposium on Usable Privacy and Security),
Santa Clara, CA, United States. August 11, 2019 - August 13, 2019.

Security Responsibility

Personal responsibility: *"I'd like to see the vendors take more responsibility and take more action to secure their own devices. But because they don't always do that and I don't always necessarily trust them to do that, I take it upon myself to be responsible for the security of these systems."* (P15_A)

Manufacturer responsibility: *"They are the prime people who are responsible for things they're making because we're not putting all the time, and energy, and money on building that stuff. So, we really don't know what is inside of this."* (P9_A)

Shared responsibility: *"If you have stronger security features that the device offers the user doesn't use, that's kind of the user's fault. If it doesn't offer certain level of security, that's the manufacturer's fault."* (P10_A)

model the U.S. might consider smart home devices.

We found disparities between privacy responsibility, concern, and mitigations. For example, P6_U and P8_A said they were at least partially responsible for the privacy of their devices but were not personally concerned. Not surprisingly, P6_U did not perform any privacy mitigations. While P8_A did cover the cameras on some of his devices, he did so only due to prompting by his wife.

Personal responsibility appears to be a differentiator when it comes to privacy concern and mitigation. Of the eight participants who said that they were at least partially responsible for privacy, seven mentioned at least one mitigation they perform. Of the six who claimed no personal responsibility, only three discussed performing a privacy mitigation.

Responsibility for Security

When asked about responsibility for the security of their smart home devices, similar to privacy, participants mentioned themselves and manufacturers, but only one said government. Eight viewed responsibility as being shared.

Nine claimed that they had at least partial responsibility, with three of those taking sole responsibility. One smart home owner remarked, *"I think we've realized, sooner or later, your stuff will get breached. It's on you to either put extra restrictions in place or just be okay with the fact that it's going to happen" (P8_A).*

Nine participants said manufacturers have at least some responsibility for security, with two of those claiming the manufacturer has sole responsibility and six believing that both the manufacturer and user hold responsibility. For example, one participant remarked, *"I consider myself to be responsible for doing the best I can security-wise, but really it's the manufacturers and the people who develop the software that ultimately hold the keys to the security" (P11_A).*

Seven of the nine claiming personal responsibility discussed some kind of security mitigation. A participant who did not implement mitigations was personally concerned, but not knowledgeable enough to take action: *"It could be fairly simple to do something and protect myself, but I have no idea. . . I'm not going to educate myself on network security. . . This stuff is not my forte. I'm very accepting to the fact that it is what it is" (P8_A).*

Two participants claimed responsibility but were not personally concerned about security. Those not claiming personal responsibility were also not personally concerned about security, with only one mentioning a rudimentary mitigation (occasionally changing passwords).

Future Work and Contributions

We plan to complete the interview study, with a goal of 40 interviews total. We would like to especially recruit more non-administrative users to explore potential differences between those who install and administer the devices and those who may only use the devices. With this larger dataset, we will also continue to investigate possible relationships between privacy and security concerns, mitigation actions, and perceptions of responsibility as well as how those perceptions and experiences interplay with usability.

Future results can inform efforts to foster a sense of personal responsibility for privacy and security among smart home end users or encourage more manufacturer or government accountability in areas for which smart home users tend not to claim responsibility. By examining the sophistication of mitigations, we can also start to uncover areas ripe for improved usable privacy and security features that manufacturers can build into their devices by default, thus alleviating the need for users to take protective action.

Haney, Julie; Furman, Susanne; Theofanos, Mary; Acar Fahl, Yasemin.
"Perceptions of Smart Home Privacy and Security Responsibility, Concerns, and Mitigations."
Paper presented at WSIW 2019 - 5th Workshop on Security Information Workers (held at the 15th Symposium on Usable Privacy and Security),
Santa Clara, CA, United States. August 11, 2019 - August 13, 2019.

Disclaimer

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

REFERENCES

1. Bill Ablondi. 2019. Connected Consumer: Chairpersons Opening Remarks. Strategy Analytics. Presented at the 2019 Internet of Things World Conference, Santa Clara, CA.
2. Parks Associates. 2019. State of the Market> Smart Home and Connected Entertainment. (jan 2019). Retrieved May 29, 2019 from http://www.parksassociates.com/bento/shop/whitepapers/files/ParksAssoc-OpenHouseOverview2018.pdf
3. Barney G Glaser and Anselm L Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
4. PwC. 2017. Smart home, seamless life. (jan 2017). Retrieved May 29, 2019 from https://www.pwc.fr/fr/assets/files/pdf/2017/01/pwc-consumer-intelligence-series-iot-connected-home.pdf
5. Statista. 2019. Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions). (2019). Retrieved May 31, 2019 from https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
6. Meredydd Williams, Jason RC Nurse, and Sadie Creese. 2017. Privacy is the boring bit: user perceptions and behaviour in the Internet-of-Things. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 181–18109.
7. Peter Worthy, Ben Matthews, and Stephen Viller. 2016. Trust me: doubts and concerns living with the Internet of Things. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 427–434.
8. Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. 65–80.

Haney, Julie; Furman, Susanne; Theofanos, Mary; Acar Fahl, Yasemin.
"Perceptions of Smart Home Privacy and Security Responsibility, Concerns, and Mitigations."
Paper presented at WSIW 2019 - 5th Workshop on Security Information Workers (held at the 15th Symposium on Usable Privacy and Security), Santa Clara, CA, United States. August 11, 2019 - August 13, 2019.

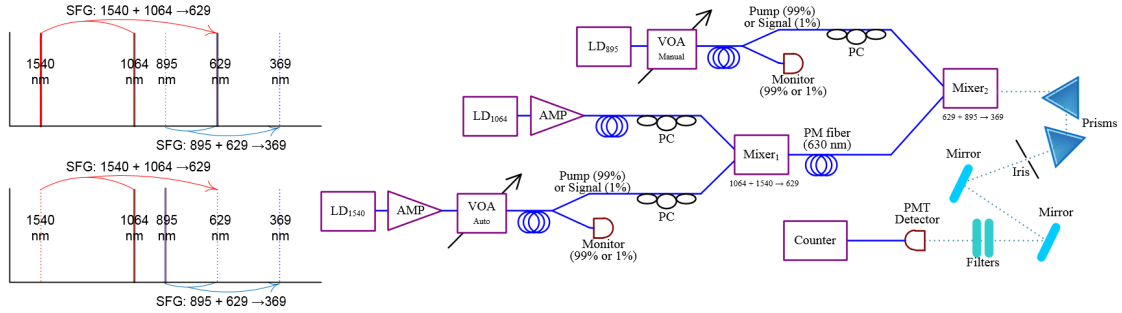
A cascaded interface to connect quantum memory, quantum computing and quantum transmission frequencies

Oliver Slattery[1,*], Lijun Ma[1] and Xiao Tang[1]
[1] *Information Technology Laboratory, National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899*
** email: oliver.slattery@nist.gov*

Abstract: We implement a cascaded interface connecting three essential frequencies for quantum communications including 1540 nm for long-distance transmission, 895 nm for Cesium quantum memory and 369 nm for Ytterbium ion quantum computing applications.
OCIS codes: (270.5565) Quantum Communications; (190.0190) Nonlinear optics; (270.0270) Quantum Optics.

1. Introduction

Future distributed quantum computing networks will require different technologies and materials to implement particular processes such as the quantum transmission, storage and computation. Typically, the most suitable material or technology varies from process to process and with it, the particular wavelength at which that process can be implemented most optimally. Quantum interfacing is therefore an essential link to enable these different processes at different wavelengths to be combined in a single network. In particular, the wavelength of 1540 nm is suitable for the long-distance transmission of quantum state encoded single photons in standard telecom optical fibers. The wavelength of 895 nm can be used for quantum state storage and quantum memory based on Cesium atoms [1]. The wavelength of 369 nm can be used for quantum computing applications based on Ytterbium trapped-ions [2]. In this paper, we introduce a cascaded interface that connects all three of the wavelengths required for these processes in a single experimental set-up. Single photons form either a Cesium based quantum memory device at 895 nm or transmitting through a fiber at 1540 nm can be selectively converted to 369 nm for use in quantum computing. The cascaded quantum frequency conversion mechanisms are illustrated in Fig. 1(a). The experimental set-up is shown in Fig. 1(b) and described in the text.



(a) (b)

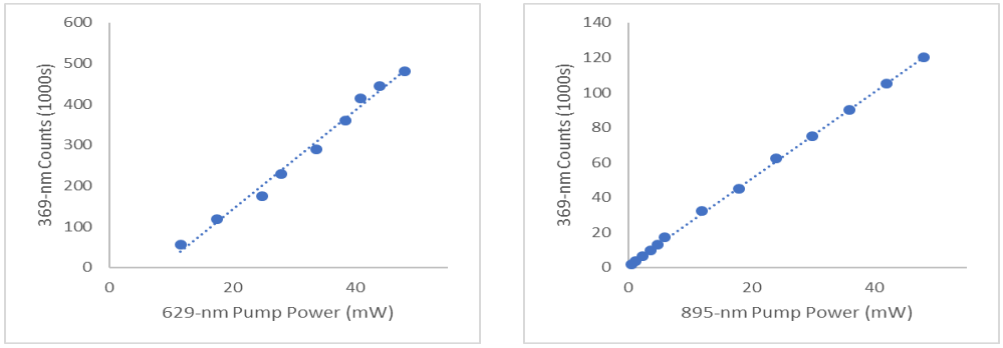
Fig. 1. (a) Illustration of the frequency conversion processes. Upper: signal at 895 nm converted to 369 nm with pump at 629 nm generated from a strong 1540 nm and 1064 nm pre-pumps. Lower: signal at 1540 nm converted to 629 nm and subsequently to 369 nm with a pump at 895 nm. SFG: sum-frequency-generation. (b) Schematic of the cascaded interface experimental set-up. $LD_{wavelength}$: laser at specified wavelength; AMP: amplifier; VOA: variable optical attenuator (manual or automatic); PC: polarization control; PMT: photomultiplier tube.

The mixers are based on sum-frequency-generation (SFG). The first mixer (SFG: 1540 nm + 1064 nm → 629 nm) has a normalized conversion efficiency of approximately 70%/W. The second mixer (SFG: 895 nm + 629 nm → 369 nm) has a normalized conversion efficiency of approximately 10%/W. Strong 1540 nm and 1064 nm pumps can be combined in the first conversion device to form a strong pump at 629 nm which can then be used to convert a single photon signal at 895 nm to 369 nm in the second conversion device. On the other hand, a single-photon signal at 1540 nm can be combined with a strong pump at 1064 nm to be converted to the intermediate wavelength of 629 nm in the first conversion device. These newly generated 629 nm single photons can subsequently be combined with the strong pump at 895 nm and converted to 369 nm in a second conversion device. In each case, the single photon

Slattery, Oliver; Ma, Lijun; Tang, Xiao.
"A cascaded interface to connect quantum memory, quantum computing and quantum transmission frequencies."
Paper presented at OSA Frontiers in Optics + Laser Science 2019, Washington, D.C., United States. September 15, 2019 - September 19, 2019.

signals are provided by a greatly attenuated signal laser. After the second mixer, the free-space output is separated from the residual pump light using two prims, an iris and a 369 nm filter. The prism legs are anti-reflection coated at 369 nm. The filter has approximately 30% transmission from 365 nm to 375 nm and provides over 80 dB attenuation otherwise. The signal is then guided into a photomultiplier tube whose detection efficiency at 369 nm is approximately 2% and whose dark-count rate is less than 200 counts per second (cps). The electrical output from the PMT goes directly to a counter. Both of these processes have been implemented and the results reported here are limited only by the amount of pump power available from our equipment.

2. Results and Discussion



(a) (b)

Fig. 2. Measured counts as a function of pump power for: (a) a 0.0035 μW 895 nm input to $Mixer_1$. (b) a 0.005 μW 1550 nm input to $Mixer_2$. The detection efficiency (detailed in the text) is primarily limited by the efficiency of the detector and the maximum pump power available for each conversion process. The continuing linear increase in detection efficiency indicates that a higher pump power will increase the overall conversion rate of the cascaded interface.

The maximum pump power available at 629 nm (from our 1064 nm and 1550 nm amplifiers) is 50 mW which gives a total conversion efficiency of the second mixer of approximately 0.5%. The total detection rate of the system is therefore approximately $3x10^{-5}$ including the filter and detector. We confirmed this detection rate by using a greatly attenuated laser at 895 nm as an input to this mixer. The detection efficiency of the system remains linearly increasing with the pump power beyond our maximum available pump (see Fig. 2(a)). A higher pump will lead to a higher conversion efficiency. We additionally performed the cascaded conversion of low-light from 1540 nm to 629 nm to 369 nm. The maximum power available from the 1064 nm laser amplifier (120 mW) gives a conversion efficiency of approximately 9% in the first mixer. A maximum pump power of approximately 50 mW at 895 nm was provided to the second mixer. The total conversion and detection rate of this scheme is therefore estimated to be $2.7x10^{-6}$. Again, we confirmed this detection rate by using a greatly attenuated 1550 nm laser as the input to the system. The efficiency remains linearly increasing with pump power (Fig. 2 (b)) and more pump power will result in greater overall conversion efficiencies.

It should be noted that the noise counts from the strong pump powers are not noticeable. Once any of the inputs required for the generation of 369 nm photons is turned off and all the other input are at their maximum available powers, the dark count rate reverts to under 200 cps – the same as when everything is turned off.

This cascaded interface will be very useful for hybrid long-distance fiber networks for distributed quantum computing that use Cesium-based quantum memories and Ytterbium-based quantum computers.

3. References

[1] L. Ma, O. Slattery, and X. Tang, "Optical quantum memory based on electromagnetically induced transparency," *Journal of Optics,* vol. 19, no. 4, p. 043001, 2017.

[2] C. Monroe and J. Kim, "Scaling the ion trap quantum processor," *Science,* vol. 339, no. 6124, pp. 1164-1169, 2013.

Slattery, Oliver; Ma, Lijun; Tang, Xiao.
"A cascaded interface to connect quantum memory, quantum computing and quantum transmission frequencies."
Paper presented at OSA Frontiers in Optics + Laser Science 2019, Washington, D.C., United States. September 15, 2019 - September 19, 2019.