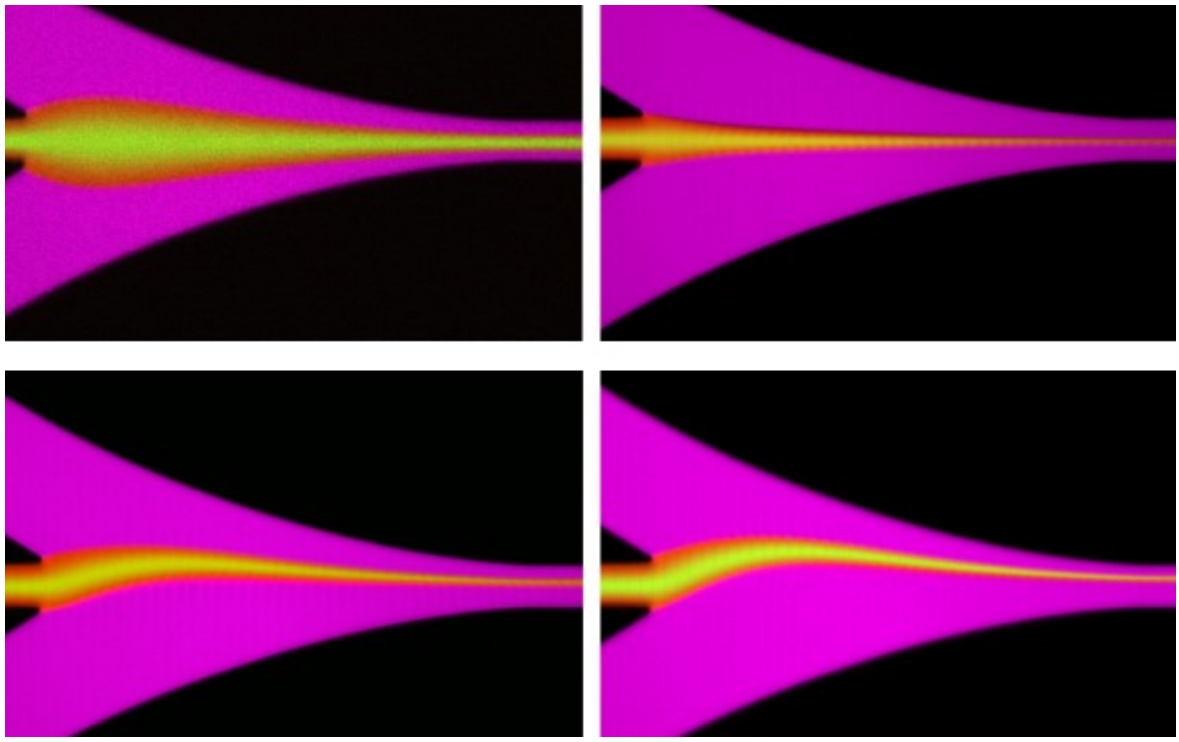
NISTIR 8373

Applied and Computational Mathematics Division

Summary of Activities for Fiscal Year 2020



This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8373

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

NISTIR 8373

Applied and Computational Mathematics Division

Summary of Activities for Fiscal Year 2020

Ronald F. Boisvert, Editor
Applied and Computational Mathematics Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8373

May 2021

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

Abstract

This report summarizes recent technical work of the Applied and Computational Sciences Division of the Information Technology Laboratory at the National Institute of Standards and Technology (NIST). Part I (Overview) provides a high-level overview of the Division's activities, including highlights of technical accomplishments during the previous year. Part II (Features) provides further details on projects of particular note this year. This is followed in Part III (Project Summaries) by brief synopses of all technical projects active during the past year. Part IV (Activity Data) provides listings of publications, technical talks, and other professional activities in which Division staff members have participated. The reporting period covered by this document is October 2019 through December 2020.

For further information, contact Ronald F. Boisvert, 100 Bureau Drive, Mail Stop 8910, NIST, Gaithersburg, MD 20899-8910, phone 301-975-3812, email boisvert@nist.gov, or see the Division's Web site at https://www.nist.gov/itl/math/.

Keywords: applied mathematics; computational science and engineering; high-performance computing; mathematics of metrology; mathematics of biotechnology; materials modeling and simulation; mathematical knowledge management; mathematical modeling; network science; scientific visualization; quantum information science.

Cover Visualization: Composite false-color, top-down views of hydrodynamic focusing in a microfluidic chip. In each panel, a core stream (red, yellow, and green) flowing left to right is compressed by two focusing flows (magenta) that impinge on it. In the top-right panel, the focusing flows have greater pressure than in the top-left, thereby creating a more tightly focused core stream. In the bottom panels, the relative pressure of the bottom focusing stream is varied to shift the core off of the centerline. The composites were generated from epifluorescence microscope images of streams spiked with fluorescent dyes. See page 52 for details. Image courtesy of Matt DiSalvo, a NIST PREP postdoctoral associate from Johns Hopkins University.

Section Visualizations: The "word cloud," which is found at the start of each Part of this document was created using Wordle, and the text of this document as input.

Acknowledgements: Thanks to Daniel Flynn for assisting in the compilation of Part III of this document and to Lochi Orr for assisting in the compilation of Part IV. Thanks also to Brian Cloteaux and Craig Greenberg who read the manuscript and offered corrections and suggestions for improvement.

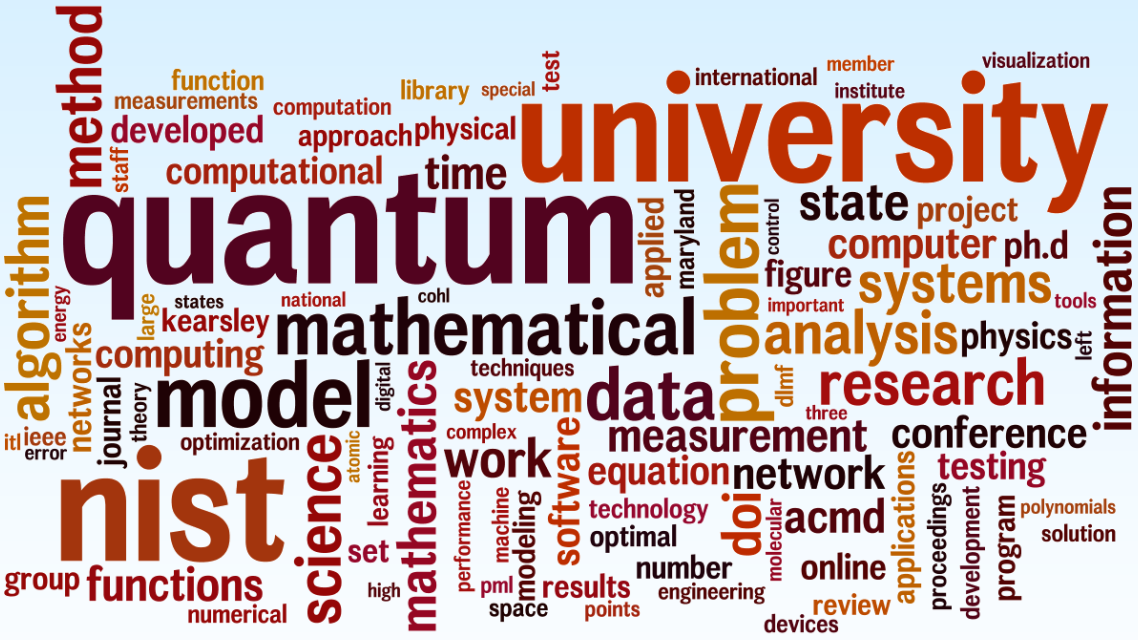
National Institute of Standards and Technology Internal Report 8373
Natl. Inst. Stand. Technol. Intern. Rep. 8373, 172 pages (May 2021)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8373

Contents

Part I

Overview



Introduction

Founded in 1901, the National Institute of Standards and Technology (NIST) is a non-regulatory federal agency within the U.S. Department of Commerce. Its mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life. The technical disciplines represented in the NIST Laboratories include physics, electrical engineering, nanotechnology, materials science, chemistry, biotechnology, manufacturing and construction engineering, fire research, information technology, mathematics and statistics. The NIST Labs operate in two locations: Gaithersburg, MD, (headquarters—234 hectare/578-acre campus) and Boulder, CO (84 hectare/208-acre campus). NIST employs about 3 400 scientists, engineers, technicians, and support personnel. NIST also hosts about 3 500 associates from academia, industry, and other government agencies, who collaborate with NIST staff and access its facilities.

The Information Technology Laboratory (ITL) is one of six major organizational units that make up the NIST Labs. ITL's singular purpose is to cultivate trust in information technology and metrology. This is done through the development of measurements, tests, and guidance to support innovation in and deployment of information technology by industry and government, as well as through the application of advanced mathematics, statistics, and computer science to help ensure the quality of measurement science.

The Applied and Computational Mathematics Division (ACMD) is one of seven technical Divisions in ITL. At its core, ACMD's purpose is to nurture trust in metrology and scientific computing. To do so, ACMD provides leadership within NIST in the use of applied and computational mathematics to solve technical problems arising in measurement science and related applications. In that role staff members

- perform research in applied mathematics and computational science and engineering, including analytical and numerical methods, high-performance computing, and visualization;
- perform applied research in computer science and engineering for future computing and communications technologies;
- engage in peer-to-peer collaborations to apply mathematical and computational techniques and tools to NIST problems;
- develop and disseminate mathematical reference data, software, and related tools; and
- work with internal and external groups to develop standards, tests, reference implementations, and other measurement technologies for current and future scientific computing systems.

Division staff is organized into four groups:

- Mathematical Analysis and Modeling Group (*Timothy Burns, Leader*). Performs research and maintains expertise in applied mathematics, mathematical modeling, and numerical analysis for application to measurement science.
- Mathematical Software Group (*Michael Donahue, Leader*). Performs research and maintains expertise in the methodology and application of mathematical algorithms and software in support of computational science within NIST as well as in industry and academia.
- High Performance Computing and Visualization Group (*Judith Terrill, Leader*). Performs research and maintains expertise in the methodologies and tools of high-performance scientific computing and visualization for use in measurement science.
- Computing and Communications Theory Group (*Ronald Boisvert, Acting Leader; Oliver Slattery, Project Leader*). Performs research and maintains expertise in the fundamental mathematics, physics, computer science, and measurement science necessary to enable the development and analysis of current and future computing and communications systems.

The technical work of the Division is organized into seven thematic areas; these are described in the sidebar. Project descriptions in Part III of this document are organized according to these broad themes.

Division Thematic Areas

Broad Areas

Mathematics of Metrology. Mathematics plays an important role in measurement science. Mathematical models are needed to understand how to design effective measurement systems and to analyze the results they produce. Mathematical techniques are used to develop and analyze idealized models of physical phenomena to be measured, and mathematical algorithms are necessary to find optimal system parameters. Mathematical and statistical techniques are needed to transform measured data into useful information. We develop fundamental mathematical methods and tools necessary for NIST to remain a world-class metrology institute, and to apply these to measurement science problems.

High Performance Computing and Visualization. Computational capability continues to advance rapidly, enabling modeling and simulation to be done with greatly increased fidelity. Doing so often requires computing resources well beyond what is available on the desktop. Developing software that makes effective use of such high-performance computing platforms remains very challenging, requiring expertise that application scientists rarely have. We maintain such expertise for application to NIST problems. Such computations, as well as modern experiments, typically produce large volumes of data, which cannot be readily comprehended. We are developing the infrastructure necessary for advanced interactive, quantitative visualization and analysis of scientific data, including the use of 3D immersive environments, and applying the resulting tools to NIST problems.

Current Focus Areas

Mathematics of Biotechnology. As proof-of-concept academic work in engineering biology meets the market realities of bringing lab science to product initiation, there are needs to compare biological products, measure whether desired outcomes are realized, and optimize biological systems for desired behaviors. NIST is working to deliver tools and standards to measure such biological technologies, outputs, and processes from healthcare to manufacturing and beyond. We support this effort with the development and deployment of innovative mathematical modeling and data analysis techniques and tools.

Materials Modeling. Mathematical modeling, computational simulation, and data analytics are key enablers of emerging manufacturing technologies. The Materials Genome Initiative (MGI), an interagency program with the goal of significantly reducing the time from discovery to commercial deployment of new materials using modeling, simulation, and informatics, is a case in point. To support the NIST role in the MGI, we develop and assesses modeling and simulation techniques and tools, with emphasis on uncertainty quantification, and collaborate with other NIST Laboratories in their efforts to develop the measurement science infrastructure needed by the materials science and engineering community.

Quantum Information Science. An emerging discipline at the intersection of physics and computer science, quantum information science is likely to revolutionize 21st century science and technology in the same way that lasers, electronics, and computers did in the 20th century. By encoding information into quantum states of matter, one can, in theory, enable phenomenal increases in information storage and processing capability. At the same time, such computers would threaten the public-key infrastructure that secures all of electronic commerce. Although many of the necessary physical manipulations of quantum states have been demonstrated experimentally, scaling these up to enable fully capable quantum computers and networks remains a grand challenge. We engage in (a) theoretical studies to understand the power of quantum computing, (b) collaborative efforts with the multi-laboratory experimental quantum science program at NIST to characterize and benchmark specific physical realizations of quantum information processing, and (c) demonstration and assessment of technologies for quantum networking.

Foundations of Measurement Science for Information Systems. ITL assumes primary responsibility within NIST for the development of measurement science infrastructure and related standards for IT and its applications. ACMD develops the mathematical foundations for such work. This can be very challenging. For example, many large-scale information-centric systems can be characterized as an interconnection of many independently operating components (e.g., software systems, communication networks, the power grid, transportation systems, financial systems). Exactly how the structure of such large-scale interconnected systems and the local dynamics of its components leads to system-level behavior is only weakly understood. This inability to predict the systemic risk inherent in system design leaves us open to unrealized potential to improve systems or to avoid potentially devastating failures. A looming new example of importance to NIST is the Internet of Things. We are developing models to aid in the development of applications from individualized health IT devices to large-scale sensor networks.

Mathematical Knowledge Management. We work with researchers in academia and industry to develop technologies, tools, and standards for representation, exchange, and use of mathematical data. Of particular concern are semantic-based representations which can provide the basis for interoperability of mathematical information processing systems. We apply these representations to the development and dissemination of reference data for applied mathematics. The centerpiece of this effort is the Digital Library of Mathematical Functions, a freely available interactive and richly linked online resource, providing essential information on the properties of the special functions of applied mathematics, the foundation of mathematical modeling in all of science and engineering.

Highlights

In this section we identify some of the major accomplishments of the Division during the past year. We also provide news related to ACMD staff. The COVID-19 pandemic provided a wealth of challenges to overcome during 2020. By mid-March the NIST campus was closed to all Division staff. A good deal of effort was expended getting those who were not already telework-ready up-to-speed. While much of our work could go on unabated, some things suffered. Our quantum optical experiments were suspended. The daily face-to-face interactions which are the lifeblood of a research organization moved online, which made them difficult at best. The spontaneity that feeds collaboration was largely lost. While staffing levels grew during the period, it became much more difficult to integrate new staff members into the life of the Division. We managed to cope with these challenges and developed new methods of interaction. By mid-summer our experimentalists, and some of our theorists who collaborate closely with experimentalists in other NIST Labs were allowed back on campus. But as I write this in spring 2021 most of our staff remain in full-time teleworking mode. We look forward to a return to normalcy later this calendar year.

Recent Technical Highlights

ACMD has made significant technical progress on many fronts during the past year. Here we highlight a few notable technical accomplishments. Further details are provided in Part II (Features) and Part III (Project Summaries).

Mathematics of Metrology. Gas chromatography, in combination with mass spectrometry, is a powerful experimental tool to identify unknown chemical compounds. The flow of the molecules in a gas chromatographic column is characterized by a dimensionless quantity called the *Kováts retention index* (KRI), which is independent of many experimental factors. NIST maintains a database of KRI of different molecules and has developed a group-contribution approach to predict KRI for other molecules. ACMD researchers have taken a novel alternative approach to that prediction, using a deep learning model based on a graph neural network which encodes elementary atomic and molecular information. The new method improves the predictive capability over the group-contribution model by almost a factor of four. See page 20.

We have taken on a leadership role in promoting the *exchange and reuse of mathematical modeling software within the atomic and molecular physics community*. A prototype gateway for the distribution and maintenance of such codes has been developed. In December 2019 ACMD hosted a workshop with software developers from that community to chart a path forward. See page 35.

High Performance Computing and Visualization. ACMD's High Performance Computing and Visualization Group has operated an *immersive visualization environment* (IVE) for many years and have successfully used it to provide critical insights to a wide variety of NIST metrology projects. To take advantage of recent advances in visualization hardware, we are moving our IVE to ParaView, a fully open source software environment. As part of this effort we are partnering with Kitware to extend and enhance ParaView's capabilities for immersive visualization. We have also joined the Khronos Group, which is working to develop standards for extended reality and for rendering scientific data. See page 86. This work has been greatly enhanced by our engagement with William Sherman, an ACMD guest researcher who, with four published books on virtual reality under his belt[1], is an acknowledged expert in this rapidly advancing field.

This year ACMD staff took a leadership role in high-performance computing at the national level through B. Schneider's service as Co-Chair of the White House's National Science and Technology Council *Fast Track Action Committee on Strategic Computing*. The group led an effort with input from government,

[1] https://www.amazon.com/-/e/B001K8O3WM

industrial, and academic stakeholders to update the goals and strategies behind the National Strategic Computing Initiative. Their report was issued in November 2019.[2]

Mathematics of Biotechnology. The on-going COVID-19 health crisis has highlighted the critical need for advanced metrology tools that can be used in both clinical and research settings. In response to the crisis, several ACMD staff members have been collaborating with MML staff on a range of projects related to the mathematics of metrology for *quantitative polymerase chain-reaction (qPCR) DNA measurements and serology-based antibody assays*. The novel data analysis techniques that they have developed can enable significant improvements in tracking the spread of COVID-19 and other infectious diseases. The generality of their mathematical analysis provides a robust framework for improving measurement accuracy because it makes so few assumptions and is widely applicable. See page 25.

ACMD staff members have made great strides in improving the rigor and fidelity of mathematical modeling in cryobiology. A focus of ACMD work has been on *cryopreservation*, the practice of freezing biological specimens, which plays an enormous role in a wide range of fields from medicine to agriculture to forensics. This period saw the publication of three papers which provide foundational models for the chemical thermodynamics, cell membrane dynamics, and solid-liquid phase transitions necessary for the study of optimization of cryopreservation protocols. See page 54.

Materials Modeling. We continue to make a positive impact to the micromagnetics modeling community through our OOMMF project, which has developed an open-source public code for *micromagnetics modeling*. OOMMF is an extensible platform in which numerical methods for micromagnetics problems can be studied. But, as a fully functional modeling system it has welcomed many users. In calendar 2020 more than 210 peer-reviewed journal articles cited use of OOMMF. To further serve the user community this year, M. Donahue and D. Porter of ACMD presented a four-session (eight-hour total) tutorial series as part of the Online Spintronics Seminar sponsored by the IEEE Magnetics Society. In addition, Donahue served as Editor of a newly published book from World Scientific Publishers entitled *Electrostatic and Magnetic Phenomena.* The volume contains a chapter contributed by Donahue and Porter on standard problems in micromagnetics. See page 67.

Quantum Information. Quantum devices are now large enough to demonstrate quantum advantage over classical computing, but it is still an open question what the first *useful* quantum task will be. Near-term algorithms need to be resistant to noise and scalable. Analog quantum algorithms, which can be used for combinatorial optimization, satisfy both these requirements, using continuous quantum evolution rather than discretized circuits. The design of efficient analog algorithms remains challenging, however. Smooth monotonic Quantum Annealing (QA) and the pulsed Quantum Approximate Optimization Algorithm (QAOA) sit on two ends of the design spectrum. We have developed a unified framework in which to treat all analog quantum algorithms and have identified conditions for *optimal analog protocols*. The resulting protocol is a hybrid algorithm, exhibiting pulse behavior at the beginning and end but often smooth, non-monotonic behavior in the middle. See page 17.

The availability of Noisy Intermediate Scale Quantum (NISQ) processors has led to attempts to describe what quantum algorithms might reasonably be implemented with low-depth quantum circuits. Jozsa has conjectured that any polynomial-time quantum computation can be simulated by a polynomial-size classical computation interleaved with polylogarithmic-depth quantum computation, so-called *hybrid quantum computation*. While this conjecture seems to hold for some important quantum algorithms (e.g., Shor's factoring algorithm), we have constructed an Oracle problem that can be solved by a polynomial-time quantum computation but cannot be solved by hybrid quantum computation. This resolves in the affirmative a conjecture due to Aaronson from 2005, and also provides the first concrete candidate for a counterexample to Jozsa's conjecture. See page 92.

Electrons confined in arrays of semiconductor nanostructures, called quantum dots, are gaining popularity as candidate building blocks for solid-state quantum devices. In semiconductor quantum computing,

[2] *National Strategic Computing Update: Pioneering the Future of Computing*. National Science and Technology Council, November 2019. URL: https://www.nitrd.gov/pubs/National-Strategic-Computing-Initiative-Update-2019.pdf

devices now have tens of individual electrostatic and dynamical gate voltages that must be carefully set to isolate the system to the single electron regime and to realize good qubit performance. Configuring these manually quickly becomes infeasible as the number of qubits grows. We are using machine learning to develop an approach to *automatically tuning quantum dots for computing applications*. Our procedure was recently successfully implemented on a double dot in collaboration with the University of Wisconsin. The article reporting this was identified as an Editor's Selection for *Physical Review Applied*. We are participating with NIST PML in a CRADA with Intel, which is interested in using this technology. See page 101.

This year our quantum communications experimental team refocused their efforts to respond to NIST's new thrust in *quantum networking*. The main tasks for this project will be to identify the key performance parameters in quantum networks; to develop versatile tools and measurement procedures for quantum networks and quantum network components; and to collaborate within and beyond NIST to support the implementation of quantum network emulation, control, management and operation. This effort has been enhanced by the recruitment of a new senior experimentalist for the ACMD program. See page 103.

Foundations of Measurement Science for Information Systems. For quite some time ACMD, in collaboration with the NIST Computer Security Division, and external organizations such as the University of Texas at Arlington and SBA Research (Austria) have been studying *combinatorial methods for software testing*. These efforts have sparked a surge of research and application of combinatorial testing (CT) technology worldwide. During this reporting period, ACMD researchers were co-authors on 10 research papers in this area. One of particular note considered the effectiveness of CT in real industrial settings. To assess this, our researchers worked with four companies to evaluate the effectiveness of CT on five industrial software systems. By comparing the results of applying CT with the internal team's own testing methods, they were able to see clear benefits of CT, especially in detecting multi-factor faults. The study also revealed some practical challenges for applying CT which will be the subject of future work. See page 114.

Mathematical Knowledge Management. This year, ACMD researchers made important strides in the use of machine learning for the *automated processing of technical documents*. Several important datasets have been developed for the community. ACMD's LaTeXML tool has been used to convert 96 % of arXiv's 1.6M documents into HTML+MathML format with a 70 % success rate. Such a resource is welcomed by the math knowledge manage research community. As one example, ACMD researchers and collaborators were able to perform a substantial study of automated methods for classifying paragraphs (according to their purpose) in technical documents. See page 125. In separate work, ACMD researchers turned the contents of NIST's Digital Library of Mathematical Functions into a well-structured, one-of-a-kind, labeled (i.e., annotated) dataset for training and testing ML models for a number of valuable math-language processing applications, and made that dataset available in the public domain. See page 125.

Technology Transfer and Community Engagement

The volume of technical output of ACMD remains high. During the last 15 months, Division staff members were (co-)authors of 58 articles appearing in peer-reviewed journals, 25 papers in conference proceedings, and nine published in other venues. Five additional papers were accepted for publication, while 14 others are undergoing review. Our staff were authors or editors of four published books. Division staff gave 33 invited technical talks and presented 13 others in conferences and workshops. Staff members were co-inventors on five new patent applications; two additional ones were awarded.

ACMD continues to maintain an active website with a variety of information and services, most notably the Digital Library of Mathematical Functions, though legacy services that are no longer actively developed, like the Guide to Available Mathematical Software, the Matrix Market, and the SciMark Java benchmark still see significant use. During calendar year (CY) 2020, the division web server satisfied more than 5.2 million requests for pages during more than 748,000 user visits. Another indication of the successful transfer of our technology is references to our software in refereed journal articles. For example, our software system for nano-magnetic modeling (OOMMF) was cited in 210 such papers published in CY 2020 alone.

Members of the Division are also active in professional circles. Staff members hold a total of 14 editorial positions in peer-reviewed journals. For example, Barry Schneider is an Associate Editor-in-Chief for IEEE's *Computing in Science and Engineering.* Staff members are also active in conference organization, serving on 23 organizing/steering/program committees. Of note, ACMD played an important role as sponsor or (co-)organizer of several significant events this year, including the following:

- *Computational Reproducibility at Exascale*[3], at SC19, Denver, CO, November 17, 2019. (M. Mascagni and Walid Keyrouz, Co-Organizers)

  This workshop addressed issues of numerical reproducibility as well as approaches and best practices to sharing and running code and the reproducible dissemination of computational results. The main target was computational reproducibility in high performance computing in general, including those issues anticipated as we scale up to exascale machines in the next decade. The participants included government, academic, and industry stakeholders.

- *A Science Gateway for Atomic and Molecular Physics*[4], NIST, Gaithersburg, MD, December 11-13, 2019. (B. Schneider, Local Organizer)

  This workshop was a follow-on to an NSF supported workshop held at Harvard's Institute for Theoretical Atomic, Molecular and Optical Physics (ITAMP) on May 14-16, 2018 entitled, "Developing Flexible and Robust Software in Computational Atomic and Molecular Physics" organized by Barry Schneider (chair), Robert Forrey (Penn State), and Naduvalath Balakrishnan (UNLV). Following the workshop six of the participating research groups interested in atomic and molecular collisions and the interaction of those systems with electromagnetic radiation submitted a joint proposal to the NSF eXtreme Science and Engineering Discovery Environment (XSEDE) to build and maintain a Science Gateway devoted to the codes developed in these groups. The goal was to explore mechanisms to collectively make codes available and easier to use by the partners as well as others in the community. This workshop focused on the next steps. With the proposal granted, the group was joined by Sudhakar Pamidighantam of the XSEDE project for development of the gateway. With some of the codes being ported to various XSEDE platforms, the group decided to focus on describing the science and the computational details of such codes to a larger community with the goal of making them available and useful to others and to also invite people who have similar interests to consider more direct participation in the project.

Service within professional societies is also prevalent among our staff. For example, Bonita Saunders was elected to the Board of Trustees of the Society for Industrial and Applied Mathematics (SIAM). Staff members are also active in a variety of working groups. Ronald Boisvert and Andrew Dienstfrey serve as members of the International Federation for Information Processing (IFIP) Working Group 2.5 on Numerical Software, Donald Porter is a member of the Tcl Core Team, Bruce Miller is a member of W3C's Math Working Group, and Sandy Ressler is a member of the Web3D Consortium. Barry Schneider represents NIST on the High-End Computing (HEC) Interagency Working Group of the Federal Networking and Information Technology Research and Development (NITRD) Program. Further details can be found in Part IV of this report.

Staff News

Once again, this year ACMD experienced an unusually large number of staffing changes. Three permanent staff members and one postdoc departed. At the same time, we welcomed three new permanent staff members, six new postdocs, one temporary hire, and one new faculty appointee. Nearly every one of our new staff members were onboarded while the NIST campus was closed during the pandemic. This has made the assimilation of new staff into Division research efforts more difficult than usual. Our senior staff members

[3] http://www.cs.fsu.edu/~cre/cre-2019/index.html

[4] https://www.nist.gov/news-events/events/2019/12/science-gateway-atomic-and-molecular-physics

have been diligent in working to give our new staff members and postdocs the best experience possible under the circumstances. The pandemic also affected our student internship programs. NIST canceled its Summer Undergraduate Research Fellowship (SURF) program, as well as its Summer High School Internship Program (SHIP). Nevertheless, we were able to support the work of 19 graduate students and four undergraduate students, mostly remotely, during the course of the year. See Table 5 on page 132 for a list of our interns.

Further details on our staff changes and awards are provided below.

Arrivals

Matthew Coudron joined the ACMD Computing and Communications Theory Group as a full-time permanent staff member in November 2019. Coudron has a Ph.D. in Theoretical Computer Science from MIT, where he studied with Peter Shor. He comes to NIST after a postdoctoral stay at the Institute for Quantum Computing at the University of Waterloo. Coudron's research interests are in quantum information and computational complexity. He is a Fellow of the Joint UMD/NIST Center for Quantum Information and Computer Science (QuICS).

Zachary Grey began a two-year appointment as a NIST NRC Postdoctoral Associate at the NIST Boulder Labs in December 2019. Zach received a Ph.D. in Computational and Applied Mathematics from the Colorado School of Mines in November 2019. His research interests are in dimension reduction and uncertainty quantification. At NIST Zach is working with Andrew Dienstfrey to apply that background to large-scale data analysis problems as well as to the quantifying the reliability of artificial intelligence models.

Thomas Gerrits, an experimental physicist by training, joined ACMD in March 2020. Thomas received a Ph.D. from Radboud University Nijmegen in the Netherlands in 2004. Since that time Thomas has performed research at NIST, first as a guest researcher and then as a staff member in the NIST Physical Measurement Laboratory in Boulder. This year he started a new chapter, working with the Quantum Communications Project in ACMD. Thomas has extensive experience in quantum optics, metrology, and single photon generation and detection. He has collaborated with ACMD staff members both in Boulder and in Gaithersburg in the past. Here he will be working on quantum network measurements and standards.

Matthew (Jake) Roberts joined ACMD in Gaithersburg in May 2020 as a NIST/NRC Postdoctoral Associate. Jake received B.S. and M.A. degrees in mathematics from Western Michigan University (2010, 2013), and a Ph.D. in mathematics from Michigan Tech (2019), where his thesis adviser was Mark Gockenback. His research area is inverse problems, specifically, the study of generalized singular value expansions. At NIST he is working with Anthony Kearsley.

Danielle Middlebrooks joined ACMD in July 2020 as a NIST Fellow Postdoc. Danielle has a B.S. in Mathematics from Spelman College (2014) and a M.S. and Ph.D. from the Applied Mathematics and Statistics and Scientific Computing (AMSC) program of the University of Maryland at College Park (2017 and 2020). Her thesis research was on geometric and Markov chain-based methods for data analysis. One application was the mutation analysis of gene regulatory networks. At NIST has been working with Jeff McFadden, Paul Patrone, and Anthony Kearsley on the NIST-in-a-Drop IMS project which is developing improved technology for cytometry.

Figure 1. *ACMD welcomed six new postdocs this period. They are (clockwise from the top left in alphabetical order): Danielle Brager, Robert DeJaco, Zachary Grey, Danielle Middlebrooks, Matthew (Jake) Roberts, and Joshua Ziegler.*

Daniel Flynn joined the ACMD Boulder staff as an Administrative Assistant in June 2020. Dan has a bachelor's degree in political science from Iowa State University (2016). For the last three years he worked with the Peace Corps in Tanzania as an agriculture volunteer leader as well as a data analyst and office manager. He is an advanced Swahili speaker. At NIST he will primarily be providing administrative support for the ITL staff in Boulder, though he will also be providing help with ACMD Gaithersburg tasks.

Robert DeJaco joined ACMD in June 2020 as a NIST/NRC Postdoctoral Associate. He received a B.S. from the University of Kentucky (2014) and a Ph.D. from the University of Minnesota (2019), both in Chemical Engineering. His research was focused on the modeling of chemical separations. To the NRC he proposed a project entitled "Accelerating Process Intensification of Membranes for Carbon Capture with Predictive Modeling." At NIST he has also been working with Paul Patrone and Anthony Kearsley on the analysis of qPCR measurements.

Joshua Ziegler joined ACMD as a NIST-NRC Postdoctoral Associate in September 2020. Josh received a B.S. in physics from Cal Poly in 2015 and a Ph.D. in Physics from the University of Oregon in 2020, where he worked on engineering and characterization of single photon emitters. In ACMD he is working with his adviser Justyna Zwolak on the use of machine learning to automate the tuning of quantum dots for use in computation.

Danielle Brager joined ACMD as a NIST-NRC Postdoctoral Associate in October 2020. Danielle received a B.S in mathematics from the University of Houston in 2010, a M.S. in mathematics from Texas Southern University in 2014, and a Ph.D. in mathematical biology from Arizona State University in 2020, where she

Figure 2. *ACMD welcomed three new permanent staff members this year to bolster its growing quantum information science program. From left to right: Matthew Coudron, Victor Albert, and Thomas Gerrits.*

focused on mathematical modeling associated with understanding the eye disease retinitis pigmentosa. Here she will work with her ACMD adviser Anthony Kearsley to apply control theory to further that effort, as well as to the study of cryobiological protocols.

Victor Albert joined ACMD as a permanent staff member in November 2020. Victor received a Ph.D. in Physics from Yale in 2017 and comes to NIST from a postdoctoral stay at Caltech's Institute for Quantum Information and Matter. He will share his time between NIST Gaithersburg and the Joint UMD-NIST Center for Quantum Information and Computer Science (QuICS) at the University of Maryland where he has been designated a QuICS Fellow. Victor has wide ranging interests related to quantum information theory including qubit encodings and error correction.

Finally, ACMD welcomed a new faculty appointee in May 2020, Professor **Debasis Mitra** of Columbia University. (Faculty appointees have special part-time Federal appointments. They typically spend about 20% of their time working at NIST; they are engaged to provide access to specialized expertise.) Prof. Mitra received a Ph.D. from London University in 1967. From 1968-2013 he worked at Bell Labs, heading the Mathematics of Networks and Systems Department for 13 years and later as Director of the Mathematical Sciences Research Center. Since 2013 he has been Professor of EE at Columbia. At NIST he is working with Vladimir Marbukh on network-based cybersecurity risk metrics.

Departures

Wesley Griffin, a computer scientist in the ACMD High Performance Computing and Visualization Group since 2012, left NIST in November 2019 for work in private industry. For seven years Griffin was a key contributor to ACMD research in immersive scientific visualization for application to NIST measurement science.

Joseph Klobusicky, a NIST-NRC Postdoctoral Associate left ACMD in August 2020 to take a position as Assistant Professor in the Mathematics Department at the University of Scranton. At NIST Joe worked on asymptotics in molecular motor models.

James Sims a computational scientist in ACMD's High Performance Computing and Visualization Group, retired in September 2020. Jim spent decades at NIST serving in a variety of roles, most recently as an expert in parallel computing. In that role he worked closely with PML scientists on the modeling of the optical properties of quantum dots. Jim was also the inventor of the Hylleraas-configuration-interaction (Hy-CI) method for the computation of properties of few-electron atomic systems. Along with long-time collaborator Stanley Hagstrom, Jim was able to use this technique to perform the world's most accurate computations of many such systems. His most recent publication, a by-product of that work, describes (and makes available) a computationally fast Fortran 90+ quadruple precision portable parallel package for solving the generalized real symmetric-definite eigenvalue problem for large (80,000 × 80,000 or greater) dense matrices. Jim remains associated with ACMD as a guest researcher.

Steven Satterfield, who served as technical manager of ACMD's immersive scientific visualization lab (the CAVE), retired in December 2020. Steve's work in visualization has enabled advances in a wide variety of NIST measurement science projects, from determining the underlying mechanism of shake gels, to measuring the critical dimensions of engineered polymer scaffolds used

Figure 3. *Two long-standing Division staff members retired this period: James Sims (left) and Steve Satterfield (right). Both will continue as guest researchers.*

in tissue engineering, to revealing the internal interactions of particles in flowing cement. The latter work was honored with the NIST Judson French Award (see below). The visualization tools he helped design and create have made ACMD's CAVE an agile virtual measurement environment in which new interactive visualizations, and the insights they provide, can be spun up quickly. Steve continues to be affiliated with ACMD as a guest researcher.

Figure 4. *Three ACMD staff members attained Fellow status in professional organizations this year. Fern Hunt was named a Fellow of the Association for Women in Mathematics, Ron Boisvert was named a Fellow of the Association for Computing Machinery, and Michael Donahue was named Fellow of the Washington Academy of Sciences.*

Recognition

ACMD staff members were recognized with a variety of awards this year, including the following.

Fern Hunt, a recent ACMD retiree and a current guest researcher, was named a Fellow of the Association for Women in Mathematics (AWM) in October 2019 for "exceptional commitment to outreach and mentoring; for her sustained efforts to make the AWM organization more inclusive; for her service to higher education and government; and for inspiring those underrepresented in mathematics with her work in ergodic theory, probability, and computation."

Ronald Boisvert was named a Fellow of the Association for Computing Machinery (ACM) in December 2019 "for contributions to mathematical software and service to the community." ACM is the largest educational and professional society in the field of computing. ACM's Fellows comprise less than one percent of the Association's global membership.

Mike Donahue, Leader of ACMD's Mathematical Software group, received the 2020 Washington Academy of Sciences (WAS) Award for Excellence in Applied Mathematics. The citation reads "For excellence in applied mathematics, leading to new tools for modeling and simulation which have transformed research into nanoscale magnetic films, structures and devices." The award, which also confers the honor of being named a WAS Fellow, was conferred in an online ceremony in September 2020.

Six ACMD staff members were honored with four different DOC and NIST Awards in 2020:

- **Bonita Saunders**: NIST Diversity, Inclusion and EEO Award. First presented in 1977, this award is granted for exceptionally significant accomplishments and contributions to Equal Employment Opportunity/Diversity goals. The award consists of an engraved plaque and a $5000 honorarium. Bonita was honored for exemplary service as a role model, mentor, and tutor in support of STEM careers by women and minorities.
- **William George** and **Steve Satterfield** : NIST Judson C. French Award. First presented in 2000, this award is granted for significant improvement in products delivered directly to industry, including new or improved NIST calibration services, Standard Reference Materials (SRMs), and Standard Reference Databases. The award consists of an engraved plaque and a $5000 honorarium. Bill and Steve were honored along with Nicos Martys of the NIST Engineering Lab and Blaza Toman of the ITL Statistical Engineering

Figure 5. *Six ACMD staff members were honored with NIST-level awards in 2020. From left to right: Bonita Saunders, William George, Steve Satterfield, Emanuel Knill, Scott Glancy, and Barry Schneider.*

Division for development of a concrete rheology standard reference material (SRM), the first SRM certified through computer simulation.

- **Scott Glancy** and **Emanuel Knill:** DOC Gold Medal for their work on the team that accomplished the world's first demonstration of deterministic quantum gate teleportation between physically separated trapped ions.
- **Barry Schneider**: DOC Bronze Medal for outstanding initiative as part of the team that developed and deployed Enki, a new computational capability in support of NIST initiatives in big data and AI.

Daniel Lozier, retired staff member from ACMD (and now a guest researcher) was inducted into NIST's "wall of fame" in October 2020. The NIST Portrait Gallery of Distinguished Scientists, Engineers and Administrators honors NBS/NIST alumni for outstanding career contributions to the work of NBS/NIST. Portraits and biographies of those selected are displayed in the corridor adjacent to the NIST cafeteria at Gaithersburg, and in the Digital Portrait Gallery at NIST Gaithersburg and NIST Boulder sites. Dan was honored for his leadership of the NIST Digital Library of Mathematical Functions project as well as for his research contributions to the field of special functions.

ACMD staff also showed well in the 2020 ITL Awards program, securing four of the nine accolades:

- **Brian Cloteaux** and **Vladimir Marbukh**: Outstanding Conference Proceedings Award for their paper, "SIS Contagion Avoidance on a Network Growing by Preferential Attachment," which was published in the Proceedings of the 2nd Joint International Workshop on Graph Data management Experiences and Systems (GRADES) and Network Data Analytics (NDA).
- **Geoffrey McFadden**: Contribution for Enhancing Diversity Award for serving as an effective teacher, mentor and adviser.
- **Paul Patrone**: Outstanding Contribution Award for his ambitious and successful research agenda which resulted in, among many things, ITL's first NIST Innovations in Measurement Science award project in seven years.
- **Barry Schneider** and **Heman Gharibnejad**: Outstanding Journal Publication Award for their paper, "Numerical Methods Every Atomic and Molecular Theorist Should Know," which was published in *Nature Reviews Physics* in December 2019.

Finally, **Paul Patrone** and **Anthony Kearsley**, along with 19 colleagues in several NIST Labs, were jointly honored by the NIST Materials Measurement Lab with their 2020 Collaboration and Teamwork Award. They were cited for the rapid development of nucleic acid-based measurement methods and standards, including RGTM 10169: SARS-CoV-2 Synthetic RNA Fragments, to support detection and characterization of the SARS-CoV-2 virus in a cross-division and -OU multidisciplinary, teamwork effort. Such collaboration is part of ACMD's DNA, and so this recognition is particularly appreciated.

In Memoriam

Bert Woodard Rust, a staff member of the ITL Applied and Computation Mathematics Division from 1980 to 2017, passed away on December 19, 2020.

Figure 6. *Bert Rust, displaying his good sense of humor, poses under NIST's Newton Apple Tree.*

Bert spent his earliest days on a farm in his native Tennessee. From 1958 to 1968 he worked in the Computing Technology Center at the Oak Ridge Gaseous Diffusion Plant while attending the University of Tennessee where he received a B.S. in Engineering Physics and an M.S. in Mathematics. In 1963 he married Marilyn A. Miller, and together they had four sons, Eric, Walter, Lawrence, and Jason. From 1968 to 1971 he worked for Tennecomp Systems, Inc. while attending the University of Illinois, where he received a Ph.D. in Astronomy in 1974. Bert's thesis included the discovery of one of the most important relationships in supernova cosmology—the relation between the peak luminosity of Type Ia supernovae and their luminosity decline rate after maximum light.[5] This discovery allowed such supernovae to be used as "standard candles," which, in turn, allowed astronomers to confirm the accelerating expansion of the universe.

In 1970 he published the book *Mathematical Programming and the Numerical Solution of Linear Equations* with Walter R. Burris. The book described new mathematical programming techniques for solving ill-conditioned systems of linear equations with various kinds of errors in the right-hand side vector. The new techniques used prior knowledge about the solution to greatly reduce the size of the class of solutions which are consistent with the right-hand side errors. The methods give interval estimates for the solution, with the sizes of the intervals being determined by the sizes of the errors in the right-hand side, and the constraints imposed on the class of acceptable solutions by the a priori information. While the primary motivation for that work was the spectrum unfolding problem of experimental physics, the insights there provided ample ammunition to attack difficult inverse problems for the remainder of Bert's career.

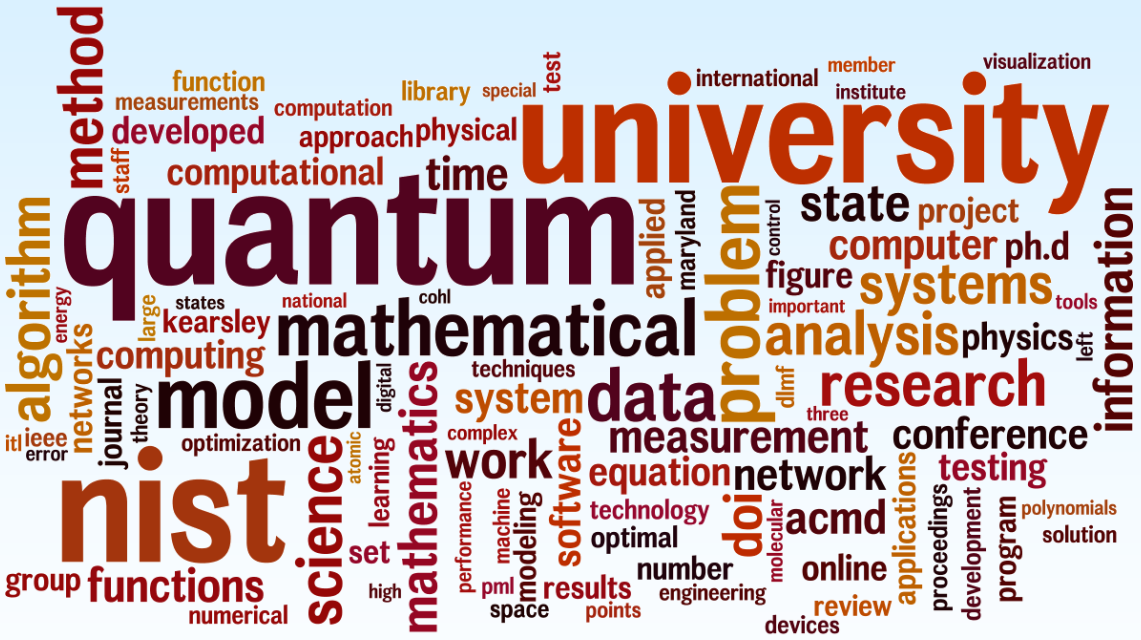
From 1972 to 1980 Bert was employed by Oak Ridge National Laboratory where he worked on environmental and ecological modeling problems. From 1980 to 2017 he was employed as a NIST mathematician. His research at NIST included ill-posed problems, time-series modeling, and nonlinear regression. He applied his techniques to a wide variety of applications, including contrast sensitivity in human vision, radar Doppler spectra, scintillation spectrometry, global warming, and observational cosmology. From 1989 to 2000, he was an adjoint professor for the Johns Hopkins University Part-Time Programs in Engineering and Applied Science where he taught graduate courses in numerical analysis and digital spectral analysis. His interest in pedagogy, was also displayed in a set of very well received tutorial articles for a series he called "Fitting Nature's Basic Functions" for *Computing in Science & Engineering* magazine from 2001 to 2003.

Bert was a true NISTer: an outstanding mathematician with many important contributions, and a pleasant, good person to boot. His legacy carries on, but he will be missed by many indeed.

[5] Maria Victorovna Pruzhinskaya and Sergey Mikhailovich Lisakov. How Supernovae Became the Basis of Observational Cosmology. *Journal of Astronomical History and Heritage* **19**:2 (2016), 203-215.

Part II

Features



Optimal Protocols for Analog Quantum Algorithms

Quantum devices are now large enough to achieve quantum advantage over classical computing, but it is still an open question what the first useful quantum task will be. Near-term algorithms need to be resistant to noise and scalable, expanding to use the available circuit depth and resources. Analog Quantum Algorithms satisfy both these requirements, using continuous quantum evolution rather than discretized circuits to achieve the power of quantum advantage, so such analog algorithms are current favorites for near-term devices.

One of the most important challenges though is how to design these analog algorithms; smooth monotonic Quantum Annealing and the pulsed Quantum Approximate Optimization Algorithm (QAOA) sit on two ends of the design spectrum for these algorithms. Optimizing these algorithms isn't just a question of efficiently using resources but can also translate directly into whether quantum advantage is achieved. Our work provided a consistent framework in which to treat all analog quantum algorithms and identified conditions for and properties of the optimal analog protocol. The resulting protocol is a hybrid algorithm, exhibiting pulse behavior at the beginning and end but often smooth, non-monotonic behavior in the middle. Further work shows that this optimal protocol even reduces to previously known algorithms in certain limits and lends itself to the creation of algorithmic procedures capable of capturing more of the optimal properties.

Lucas Brady

Analog Quantum Algorithms or Hamiltonian Algorithms utilize quantum dynamics via the Hamiltonian (energy/time-translation operator) directly to modify a system and attempt to prepare a desired quantum state. This reliance on time evolution of the entire system puts analog algorithms in contrast with digital quantum algorithms which use discretized circuits to manipulate portions of the system locally. That said, the analog paradigm possesses the full power of quantum computation and has many advantages for near term computers. Notably, many analog algorithms are resistant to noise or can naturally take noise into effect via their variational nature, and analog quantum algorithms lend themselves to a different design paradigm that focuses on physical properties of the system rather than quantum logic, making them especially suitable for tailoring to specific hardware.

Analog algorithms are exemplified by Quantum Annealing [1] and the very similar Quantum Adiabatic Optimization [2] which both rely on two Hamiltonians. The algorithms start in the ground state of one of the Hamiltonians, $\hat{H}_0$ and then ramp the system into the other Hamiltonian, $\hat{H}_1$, trying to stay in the ground state to measure the properties of the desired ground state of the final Hamiltonian,

$$\hat{H}(t) = u(t)\,\hat{H}_0 + (1-u(t))\,\hat{H}_1. \qquad (1)$$

For Quantum Annealing, $u(0) = 1, u(t_f) = 0$ at the final time t_f, and between a continuous monotonic ramp is employed. The adiabatic theorem guarantees the success of these algorithms if the runtime is long compared to the inverse gap or energy difference between the ground state and first excited state during the anneal. Furthermore, the shape of the ramp is often important to achieving a quantum advantage [3], but both the shape and the spectral gap can be impossible to know without first solving the problem of interest.

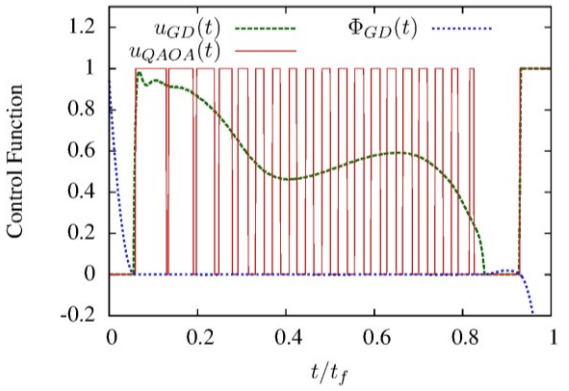


Figure 7. *Optimal control functions found through either gradient descent, $u_{GD}(t)$, or time constrained QAOA, $u_{QAOA}(t)$, for a random instance of the MaxCut problem. Also shown is the gradient $\Phi_{GD}(t)$ for the gradient descent method.*

Another important near-term analog algorithm is the Quantum Approximate Optimization Algorithm (QAOA) [4] which is closely related to the Variational Quantum Eigensolver (VQE) from quantum chemistry applications [5]. These algorithms can still be described using Eq. (1), but instead of $u(t)$ being a smooth, monotonically decreasing function, it now takes on a bang-bang form where it rapidly switches between $u = 0$ bangs and $u = 1$ bangs. There are $2p$ such bangs (p layers each consisting of a $u = 0$ bang and a $u = 1$ bang), and the lengths of these bangs are treated as variational parameters. The goal of the algorithm is then to run an optimization over these bang lengths, looking for the configuration that minimizes the energy of the output state. The optimization loop can be done on a classical computer with the quantum computer carrying out the computationally difficult task of running the procedure to determine the final state.

In practice though, many systems possess the ability to implement arbitrary $u(t) \in [0,1]$, so a major question, not only for pure efficiency but also for achieving

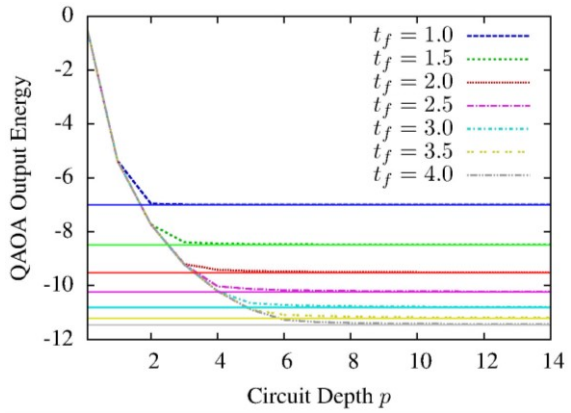


Figure 8. *QAOA output energy as a function of the circuit depth p (i.e., number of layers), for an n = 10 MaxCut instance on a 4-regular graph. Dashed lines are the QAOA energies; solid horizontal lines are the energies obtained from gradient descent.*

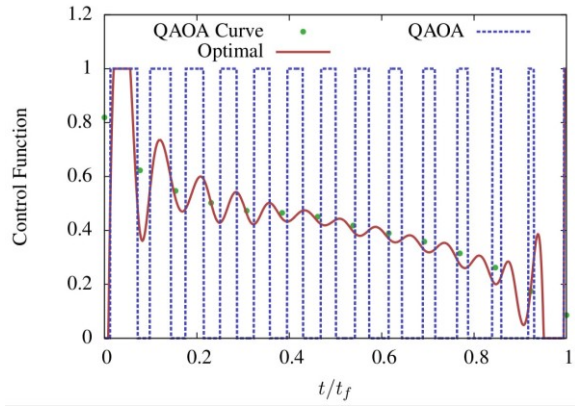


Figure 9. *An optimized QAOA protocol compared to the optimal bang-anneal-bang protocol that takes the same time. The QAOA curve is the fraction of each layer that is in a u = 1 bang and falls along a single line for a given problem independent of how many layers p are used. The number of bangs matches up with the oscillations in the optimal curve and the annealing region of the curve falls along the QAOA curve. Our work provides proofs and evidence for why these properties match up.*

quantum advantage, becomes how to design a protocol that achieves the best (i.e., lowest energy) state at the end of the evolution. Some early results [6] indicated that the optimal protocol takes on a bang-bang form based off comparison to similar classical systems. These results relied on optimal control theory [7], an extension of the Calculus of Variations, but made assumptions about the rarity of so-called singular regions that our results ultimately proved were unjustified.

Our work [8] resumes the study of this problem using optimal control theory, formalizing its usage and deriving accurate results about how the optimal protocol behaves. To explicitly state the goal, we aim to minimize the final state energy subject to evolution under the Schrödinger equation

$$\langle E \rangle = \langle x(t_f)|\hat{H}_1|x(t_f)\rangle \quad (2)$$

$$\frac{d}{dt}|x(t)\rangle = -i\hat{H}(t)|x(t)\rangle \quad (3)$$

Here $|x(t)\rangle$ is a normalized, complex vector that represents the state of the system and $\langle x(t)|$ is its Hermitian conjugate. Optimal control theory then asks what the conditions on $u(t)$ would be to ensure that $\langle E \rangle$ is as low as possible. This is carried out via the Calculus of Variations, looking for the $u(t)$ that satisfies the conditions for a critical point, subject to the condition that $u(t) \in [0,1]$. The Schrödinger differential equation is imposed via a Lagrange multiplier $|k(t)\rangle$, and it turns out that this Lagrange multiplier must also satisfy the Schrödinger equation along with the boundary condition $|k(t_f)\rangle = \hat{H}_1|x(t_f)\rangle$.

The end result of this control analysis is the gradient of the cost function with respect to the control function $\Phi(t) = \delta\langle E \rangle/\delta u(t)$. Without any constraints, $\Phi(t) = 0$ would be necessary for an optimal protocol, but because $u(t) \in [0,1]$, $\Phi(t)$ can be non-zero if $u(t)$ is on a boundary. Our analysis showed that just based on boundary conditions in the problem, $\Phi(t)$, which is continuous, must be non-zero at the beginning and end of the procedure, necessitating a $u = 0$ bang at the beginning and a $u = 1$ bang at the end. Much can be derived and proven about the lengths of these bangs as well as the form $u(t)$ must take if $\Phi(t) = 0$. Ultimately, we identified a common pattern that many optimal protocols take, which we termed *bang-anneal-bang*. Such protocols, exemplified by Figure 7, have bangs at the beginning and end but a smooth non-monotonic annealing-like curve in the middle. This pattern was extremely common for quadratic Ising/MaxCut problems, with other models exhibiting similar behavior, often with additional bangs in the middle.

Furthermore, also shown in Figure 7 is the result of a QAOA procedure constrained to take the same amount of time. Such time-constrained QAOA protocols are actually trying to emulate the optimal bang-anneal-bang protocol, digitizing the underlying annealing region. Furthermore, the energy of these protocols converges in the limit of large p, as seen in Figure 8.

These techniques do more than just find the optimal protocol. Our continuing work [9] shows even more connections between this optimal protocol and both QAOA and Quantum Annealing. To see these connections, you can plot a time-unconstrained QAOA protocol alongside the optimal bang-anneal-bang curve for that same time as in Figure 9. The number and locations of the bangs line up almost exactly with the oscillatory pattern in the underlying bang-anneal-bang curve. In our upcoming work, we show how this large-time step discretization (Product Formula approximation) occurs, finding a reduction in the error of the product formula approximation if the lengths of the bangs correspond to the underlying oscillations in the curve. Furthermore, we explore why these curves oscillate, connecting this behavior to nearly adiabatic interactions between the two lowest energy states.

Furthermore, the limit of long times, the optimal bang-anneal-bang curves smooth out, with the lengths of the initial and final bangs going to zero and the oscillation amplitudes decreasing. In that sense, these curves start approaching the smooth, monotonic behavior of Quantum Annealing curves in the adiabatic limit. Furthermore, the shape of that curve seems to mesh up with the optimized annealing curves already studied by the community [3].

In the future we hope to leverage the shape of these optimal curves to improve the performance of near-term algorithms, working with experimentalists to realize these kinds of curves in actual devices. Our initial results indicate this can lead to improvement over existing QAOA and Annealing techniques without significantly more additional resources.

References

[1] T. Kadowaki and H. Nishimori. Quantum Annealing in the Transverse Ising Model. *Physical Review E* **58** (1998), 5355. DOI: 10.1103/PhysRevE.58.5355

[2] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum Computation by Adiabatic Evolution. arXiv:0001106 (2000).

[3] J. Roland and N. J. Cerf. Quantum Search by Local Adiabatic Evolution. *Physical Review A* **65** (2002), 042308. DOI: 10.1103/PhysRevA.65.042308

[4] E. Farhi, J. Goldstone, and S. Gutman. A Quantum Approximate Optimization Algorithm. arXiv:1411.4028 (2014).

[5] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien. A Variational Eigenvalue Solver on a Photonic Quantum Processor. *Nature Communications* **5** (2014), 4213. DOI: 10.1038/ncomms5213

[6] Z. C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon. Optimizing Variational Quantum Algorithms Using Pontryagin's Minimum Principle. *Physical Review X* **7** (2017), 021027. DOI: 10.1103/PhysRevX.7.021027

[7] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The Mathematical Theory of Optimal Processes*. Interscience Publishers, New York, 1962.

[8] L. T. Brady, C. L. Baldwin, A. Bapat, Y. Kharkov, and A. V. Gorshkov. Optimal Protocols in Quantum Annealing and QAOA Problems, *Physical Review Letters*, to appear. arXiv:2003.08952 (2020).

[9] L. T. Brady, L. Kocia, P. Bienias, A. Bapat, Y. Kharkov, A. V. Gorshkov. In preparation.

Participants

Lucas Brady (ACMD); Christopher Baldwin, Alexey Gorshkov (NIST PML); Aniruddha Bapat, Yaroslav Kharkov, Przemyslaw Bienias (University of Maryland); Lucas Kocia (Sandia National Labs)

Predicting Kováts Retention Indicies Using Machine Learning

Gas chromatography, in combination with mass spectrometry, is a powerful experimental tool to identify unknown chemical compounds. The flow of the molecules in a gas chromatographic column is characterized by a dimensionless quantity called the Kováts retention index (KRI), which is independent of many experimental factors. NIST maintains a database of KRI of different molecules and has developed a group-contribution approach to predict KRI for other molecules. In the present work, we have taken a novel alternative approach to that prediction, using a deep learning model based on a graph neural network which encodes elementary atomic and molecular information. The new method has proven to be quite successful, improving the predictive capability over the group-contribution model by almost a factor of four.

Chen Qu

Gas chromatography (GC) is an important analytical technique for the separation of chemical compounds. It is frequently used in combination with mass spectrometry (GC/MS) as a means of enhancing the accuracy of the identification of chemical compounds. A GC experiment consists of putting an unknown substance in a gaseous state into a carrier gas and then passing it through a chromatography column. Interactions with compounds used in the column control the elapsed time before the compound elutes from (i.e., passes through) the column.

The retention time for the compound in the column, indexed against those of known compounds, is called the retention index. Research in the 1950s by Ervin Kováts [4] demonstrated that the retention index could be made independent of many experimental factors such as column length, column diameter, and film thickness. This produces a dimensionless quantity known as the Kováts retention index (KRI), which is the subject of the current investigation.

The identification of unknown compounds using mass spectrometry is considerably enhanced by matching the retention index against those in a library of measured compounds. This two-step process considerably improves the confidence of the identification scheme over mass spectrometry alone [3]. Thus, there is significant value in developing predictive models for the Kováts retention index for mass spectral library searching. An accurate predictive model should lead to corresponding improvements in library searches, smaller lists of possible matches with fewer false positives.

Figure 10. *Example of a gas chromatograph. (Image from Wikipedia.)*

NIST maintains a database of KRI of different molecules[6] and has developed a group-contribution approach to predict KRI [8]. While this approach has had reasonable success, our motivation for this work has been to develop a more general and accurate model by utilizing techniques from machine learning and artificial intelligence informed by chemical information.

In FY 2019, we trained a few different neural network (NN) models on the NIST 17 library[7], the latest version of the KRI data base available, including a multi-layer perceptron, SchNet, [7] and a graph NN, which made use of the MatErials Graph Network (MEGNet) model developed by Chen et al. [2]. Among these three models, the MEGNet graph NN achieved the best results and showed great potential for predicting KRIs. It produced a root-mean-square error (RMSE) of ≈72 KRI units and a mean absolute error (MAE) of ≈42 KRI units; both are significantly better than those of the group-contribution model currently used in the NIST MS library.

The MEGNet model incorporates a number of features that make it a good choice for machine learning of

[6] NIST Standard Reference Database 1A: NIST/EPA/NIH Mass Spectral Library, https://www.nist.gov/srd/nist-standard-reference-database-1a

[7] https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata%3Ari-database

molecular properties. In particular, a graph neural network (GNN) is employed [1]. The graph used by the machine learning model corresponds directly to the molecular structure. The nodes/vertices in the graph correspond to the locations of the atoms, and the edges correspond to atomic pairs (there is an edge between any pair of atoms, even if they are not formally chemically bonded). Thus, the chemical graph model may be directly mapped onto the GNN model. Properties of atoms and bonds may be assigned to the nodes and edges in the GNN model. The MEGNet framework also allows global properties (i.e., not associated with a single atom or chemical bond) of a molecule to be incorporated into the model.

The MEGNet model is thoroughly described in the literature and online, where the MEGNet Source Code[8] is also available. We only discuss the salient features of the model so that it may be reproduced by others. MEGNet version 0.3.5 is used with TensorFlow[9] version 1.14.1 and the Keras[10] API. MEGNet encodes input data in three distinct vectors (called *attributes*): state attributes, edge attributes, and node attributes. State attributes are global properties of the molecule (e.g., numbers of atoms, molecular mass). Edge attributes are features of atom pairs (e.g., bond order, member of ring). Node attributes are features of the atoms themselves (e.g., atomic number, hybridization). Each of these attributes are used to update the other attributes and are themselves updated during model fitting.

Our MEGNet model incorporates three atom features, three edge features, and three global features. Here "one-hot" encoding is extensively used in these feature vectors. For example, categorical data that have three categories can be encoded as [0, 0, 1], [0, 1, 0], and [1, 0, 0], respectively, for category 1, 2, and 3. The atom features are the atomic number (11 one-hot variables representing C, H, O, N, F, Si, P, S, Cl, Br, and I), the hybridization of the atom (six one-hot variables, s, sp, sp2, sp3, sp3d, and sp3d2), and the formal charge (one integer variable) on the atom. The edge features are the bond type (i.e., no bond, single, double, triple, or aromatic, five one-hot variables), whether the atoms were in the same ring (a single one-hot variable), and a graph distance (one integer variable; the graph distance is calculated as the smallest number of chemical bonds between the atoms in the pair).

The graph distance is calculated as the smallest number of edges that have chemical bonds between the atoms in the pair. (Recall that an edge here means any pair of two atoms, not just those that are chemically bonded). We do not consider pairs with a graph distance higher than five to save computational resources as discussed below. Global features are the number of heavy (non-hydrogen) atoms in the molecule, the molecular mass divided by the number of heavy atoms, and the number of chemical bonds divided by the number of heavy atoms.

Our MEGNet model was used with three MEGNet blocks. These blocks are composed of two densely connected layers of neurons and a graph neural network layer in which each of the attributes is successively updated. The dense layers used 128 and 64 units, respectively. The MEGNet block steps are followed by a Set2Set layer [9] in which the output of the bond and atom attributes are mapped to the appropriate vector quantities. This is followed by a concatenation step and two densely connected layers with 64 and 32 neurons.

The MEGNet source code was modified to limit the depth of atom pair properties in order to save computer memory, which scales roughly as the square of the number of atoms times the number of features. With up to 151 atoms in a molecule, the CPU memory requirements ran into the 100s of GB during model development. (By using batching, GPU memory was not a bottleneck during training.) We also implemented a custom molecule class that relied on features generated using RDKit[11].

After our initial FY 2019 efforts on the MEGNet model, we turned to the latest NIST 20 KRI Library, a much larger and more diverse data set than the NIST 17 library used in 2019. Several improvements on the previous MEGNet model were incorporated which reduced the MAE and RMSE to ~28 and 58 RI nits, respectively [6]. The improvements we made include:

- Using the latest NIST 20 MS Library. In FY 2019 we discovered a few mistakes in the NIST 17 library and removed unreliable data points from the library when training the model. ("Unreliable" was based on the magnitude of the difference between the group-contribution model and the experimental value. However, this criterion does not have a solid justification.) The NIST 20 library spans a larger chemical space, contains a more diverse set of molecules, and is of higher quality as thousands of unreliable data points were removed from the library. The following pre-processing was performed:
  a. Only data points measured on semi-standard non-polar columns in GC were used for the model, because the majority of the data were measured on this type of column.
  b. Molecules with mass less than 50 amu and more than 850 amu were removed from the set. Similarly, molecules with KRI less than 300 and more than 4100 were omitted. These extremes are not well represented in the library (only 246 molecules removed).
  c. Only molecules containing H, C, N, O, F, Si, P,

[8] https://github.com/materialsvirtuallab/megnet
[9] https://www.tensorflow.org/
[10] https://keras.io/
[11] https://www.rdkit.org/

Table 1. *Statistics of the 10-fold cross validation procedure. The mean value and standard deviation of the mean absolute error (MAE) and the root mean square error (RMSE) over 10 runs is given for each of the three sets used.*

Set	MAE	RMSE
Training	9.38 ± 0.86	18.27 ± 1.87
Validation	27.84 ± 0.67	57.77 ± 1.97
Testing	28.09 ± 0.72	58.43 ± 1.93

Table 2. *Statistics describing the graph convolutional network model performance in predicting the Kóvats retention index (RI) for the training, validation, and testing sets. The error is calculated as* $\varepsilon = RI_{\text{experiment}} - RI_{\text{predicted}}$. *The sample standard deviation, s, is calculated for the signed and unsigned errors.*

Quantity	Training	Validation	Testing
n	83 719	10 464	10 464
mean ε	–1.06	–1.23	–0.47
RMSE	20.69	54.61	57.90
min ε	–1 163	–979	–832
max ε	1 013	586	1 175
median |ε|	6.60	12.20	12.50
mean |ε|	10.20	27.69	28.67
s(|ε|)	18.03	47.09	50.30
median |%ε|	0.32	0.60	0.62
mean |%ε|	0.52	1.42	1.49
s(|%ε|)	1.00	2.56	2.84
max |%ε|	73.60	49.13	66.74

S, Cl, Br, I, were considered, as there is insufficient data for other elements. The final data set consists of 104,650 molecules, and it was randomly split into training, validation, and test subsets with ratio 8:1:1.

- Early stopping was used to mitigate over-fitting. The validation loss was monitored during the training, and if it was not improved in 150 epochs, the training was stopped, and the best model restored. Though over-fitting is still present, the early-stopping strategy at least mitigates it.
- A switch from mean-squared-error loss function (L2 norm) to mean-absolute-error (L_1 norm). The MAE significantly dropped from ≈42 to 32 with the L_1 loss function. Other loss functions, such as a weighted sum of L_1 and L_2 losses, L_∞ loss, and the Huber loss, were examined. The L_1 loss appears to be the best among all loss functions tested.
- The batch size and learning rate in stochastic gradient descent was optimized. In the 2019 work, a batch size of 128 and a learning rate of 2×10^{-3} were employed; these are currently reduced to 32 and 2×10^{-4} respectively. The MAE and RMSE were reduced to 28 and 59 RI units respectively.

The results of the 10-fold cross-validation procedure is presented in Table **1**. There is clear evidence of over-fitting in the present model as evidenced by the increase in the MAE of 19 index units and in the RMSE of 40 index units. The similar values of the validation and testing set errors are an indication that the validation set, and early-stopping are effective. Despite over-fitting, the MAE and RMSE of the testing set are acceptable and are indicative of a valid model for prediction of the KRI. The sample standard deviations of both error metrics over the 10-fold cross validation procedure indicate that the model fitting is robust.

In order to further characterize the performance of the model, the results from a single run in the cross-validation set are now analyzed in detail. These results are presented in Table 2. This run has a MAE nearly identical to the mean produced by the 10-fold cross validation and thus serves as a good representative of the model. There are several interesting details to notice in the table. The standard deviation of the absolute error (i.e., the RMSE) is consistently larger than the standard deviation of the signed error. The mean (absolute) percent error (MPE) is 1.49% for the testing set with a standard deviation of 2.84 %. The largest errors produced by the model are significant. The source of these deviations is unclear and may be due to an inadequacy of the model or an error in the source data. (Whether large prediction errors in the model correspond to errors in the source experimental data, is currently under investigation.)

Figure 11 and Figure 12 further characterize the results. Figure 11 is a histogram of the absolute errors for all three sets (training, validation, and testing). The figure clearly shows that most errors are small and the number of compounds with larger errors decreases monotonically. Figure 12 compares the predicted data and experimental values taken from the NIST 20 database. The red line shows 2 standard deviations in the MAE for each of the three sets.

The model produces a MAE of 28.09 index units and an RMSE of 58.43 index units. This is a modest improvement over the results of Matyushin et al. [5] who reported values of 33.2 and 63.0 for the MAE and RMSE, respectively. However, the present results were obtained with a data set more than three times larger than that used by Matysuhin et al. To test the two neural network models more fully, we implemented the convolutional neural network model of Matyushin et al. on the NIST 20 dataset in Python using TensorFlow. Our implementation of that model achieved an MAE of 44.4 and an RMSE of 74.6 index units in a 10-fold cross-validation test, both of which are larger than those of our graph network model. In addition, the GNN approach is a significant improvement over the model of Stein et al. [8] which shows an MAE of 114.3 and an RMSE of 166.5 (standard deviation of 121.1). The larger errors of the Stein et al. model are likely due to inadequacies in the group increment methodology and perhaps the use of the less powerful classical method (linear least squares) for model optimization. A direct comparison of the three models mentioned here is in Table 3. It presents

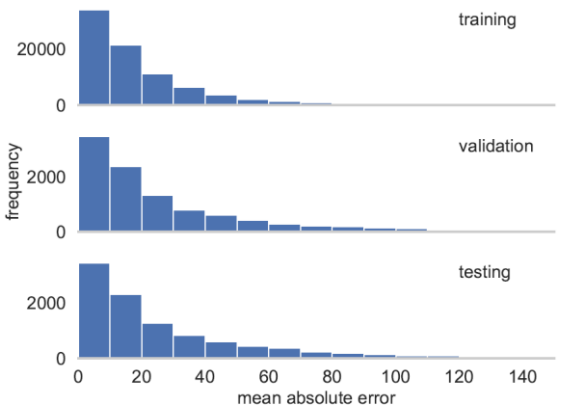


Figure 11. *Histogram of mean absolute errors in predicting the Kováts retention index (experimental – predicted) in the training, validation, and testing data sets.*

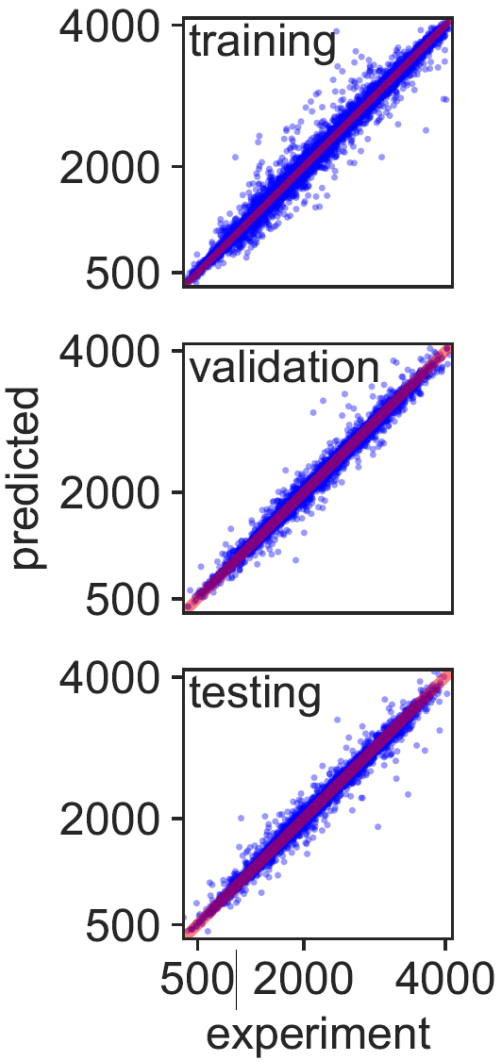


Figure 12. *Comparison of predicted to actual values of the Kováts retention index. The red line indicates perfect agreement; the width of the line equals 2 standard deviations of the prediction error.*

mean absolute errors and the sample standard deviation of the absolute mean errors for the three predictive models for the indicated chemical functionalities. In all of the cases presented, the model of Stein et al. [8] exhibits the largest error statistics of the three models. This is consistent with the conclusions drawn by Matyushin et al. [5] and from the present work. The Stein model works particularly well for hydrocarbons and exhibits the largest errors for O-atom heterocycles.

The largest errors produced by the two neural network models come from molecules with hydroxyl groups, i.e. alcohols and carboxylic acids. Performance on heterocycles, ketones, and molecules that contain S atoms also have larger MAE values as well as larger sample standard deviations. While we are not aware of any systematic bias in either model that would favor one class of compound over another, these results suggest that there may be room for improvement by explicitly considering these factors. However, this could also be a symptom of data with larger uncertainties.

The work described has demonstrated that machine learning using a GNN model is a very effective approach to the problem of predicting an experimentally derived set of Kováts retention indices. The model requires only information that can be readily obtained from a 2D representation of a molecule (i.e., a traditional drawing of a molecule). The model has been trained on a diverse set of more than 100 000 molecules, using 90 % of the data set in training and validation and using the remaining 10 % of the data set for testing the performance of the model. The model predicts the Kováts retention index with a MAE of 28 (with a corresponding RMSE of 58) versus the experimental data. The performance of the model is better than the one reported by Matyushin et al. [5] and significantly better than the widely used model of Stein et al., based on the group contribution methodology which has an MAE of 114 and a standard deviation of 167 over the same data set [8].

The success of the methodology has encouraged us to apply it to predicting the normal boiling points of compounds. The normal boiling point (BP) is the temperature at which the bulk of a liquid can change from liquid to gaseous phase. Knowledge of the BP of a compound provides crucial information such as overall volatility and purity of the system. It is regularly estimated, tabulated, and recorded in catalogs of chemical reagents. We have obtained a data set from NIST Thermodynamics Research Center[12] collection which contains the BP data of more than 4000 compounds. By applying the GNN model on this collection, we have already discovered several errors, and this opens the possibility of a research project on identifying outliers/errors in a data set.

[12] https://trc.nist.gov/

Table 3. *Evaluation of the performance of the model of Stein et al. [8], the model of Matyushin et al. [5], and the model described in this article for selected chemical functionalities for the (unitless) Kováts retention index. The column labelled N indicate the number of data used in computing the mean (μ) and sample standard deviation (s) of the absolute error,* $| | = |RI_{\text{experiment}} - RI_{\text{predicted}}|$.

		Stein		Matyushin		present work	
Molecule type	*N*	*μ*	*s*	*μ*	*s*	*μ*	*s*
all compounds	102761	113.84	119.44	38.97	57.08	27.74	50.00
ether	55755	112.60	120.21	34.66	52.16	22.99	43.87
amide	24033	119.26	112.20	38.75	56.73	25.96	47.88
contains O	92136	114.36	119.88	38.03	56.90	26.65	49.64
hydrocarbon	2010	61.67	94.10	33.27	49.57	27.77	46.75
aromatic	70501	121.93	119.82	42.48	59.36	29.68	50.95
has a ring	79091	125.44	125.90	43.37	60.77	30.80	53.19
contains N	51536	124.71	119.04	46.78	64.90	33.51	56.98
aldehyde	1176	114.23	121.06	45.89	55.80	38.66	49.55
contains S	9707	127.60	120.98	52.85	71.25	39.87	64.53
N heterocycle	22382	155.50	138.27	61.12	75.80	46.49	67.99
ketone	5611	168.16	172.95	60.98	80.86	46.94	72.89
O heterocycle	9488	182.43	173.88	63.13	73.99	48.28	66.49
alcohol	7102	141.22	153.86	64.01	80.56	52.26	75.42
carboxylic acid	1444	101.78	123.98	63.44	95.51	56.09	94.95
contains P	35	131.00	172.38	66.18	79.00	66.31	76.85
contains P	994			39.64	79.09	31.70	73.73

References

[1] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. F. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, Ç. Gülehre, H. F. Song, A. J. Ballard, J. Gilmer, G. E. Dahl, A. Vaswani, K. R. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu. Relational Inductive Biases, Deep Learning, and Graph Networks. arXiv:1806.01261, 2018.

[2] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chemistry of Materials* **31** (2019), 3564–3572.

[3] W. P. Eckel and T. Kind. Use of Boiling Point-Lee Retention Index Correlation for Rapid Review of Gas Chromatography-Mass Spectrometry Data. *Analytica Chimica Acta* **494** (2003), 235–243.

[4] E. Kováts. Gas-Chromatographische Charakterisierung Organischer Verbindungen. Teil 1: Retentionsindices Aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helvetica Chimica Acta* **41** (1958), 1915–32.

[5] D. D. Matyushin, A. Y. Sholokhova, and A. K. Buryak. A Deep Convolutional Neural Network for the Estimation of Gas Chromatographic Retention Indices. *Journal of Chromatography A* **1607** (2019), 460395–8.

[6] C. Qu, B. I. Schneider, T. Allison, A. J. Kearsley, and W. Keyrouz. Predicting Kováts Retention Indices Using Graph Neural Networks. *Journal of Chromatography A*, to appear.

[7] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A Deep Learning Architecture for Molecules and Materials. *Journal of Chemical Physics* **148** (2018), 241722–11.

[8] S. E. Stein, V. I. Babushok, R. L. Brown, and P. J. Linstrom. Estimation of Kováts Retention Indices Using Group Contributions. *Journal of Chemical Information and Modeling* **47** (2007), 975–980.

[9] O. Vinyals, S. Bengio, and M. Kudlur. Order Matters: Sequence to Sequence for Sets. arXiv:1511.06391, 2015.

Participants

Anthony Kearsley, Barry Schneider (ACMD); Walid Keyrouz (NIST ITL); Thomas Allison (NIST MML); Chen Qu (University of Maryland)

Advanced Data Analysis for Biological Measurements

The on-going COVID-19 health crisis has highlighted the critical need for advanced metrology tools that can be used in both clinical and research settings. For example, accurate diagnostic testing is needed to characterize epidemiological patterns, manage healthcare infrastructure, and identify optimal policy decisions. Likewise, development of new medical treatments requires precision measurements to determine efficacy. The implications for the current pandemic are so significant that in FY 2020 alone, Congress allocated more than $300 million to improve the quality of a single type of test for detecting SARS-CoV-2 antibodies [1].

Motivated by such issues, several staff members in ACMD have been collaborating with MML staff on a range of projects related to the mathematics of metrology for quantitative polymerase chain-reaction (qPCR) DNA measurements and serology-based antibody assays. Both techniques have enabled significant improvements in tracking the spread of COVID-19, although they are useful for detecting a much larger number of diseases as well. While these measurement techniques are nominally very different, our work highlights a broad theme that connects them: the generality of mathematical analysis provides a robust framework for improving measurement accuracy because it makes so few assumptions and is widely applicable.

Paul Patrone

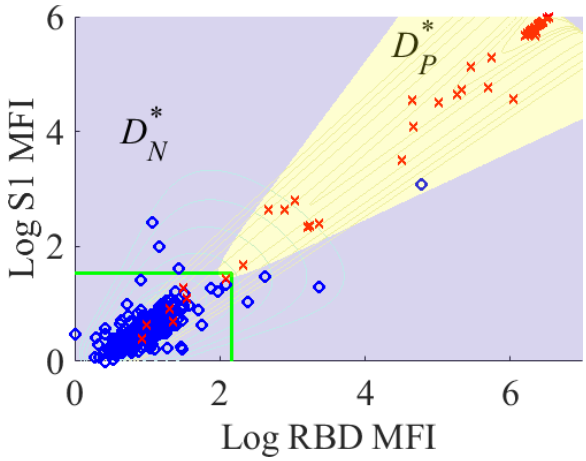


Figure 13. *Classification domains for data in [4]. The horizontal and vertical axes correspond to median fluorescence intensities (MFI) associated with antibody levels that target the receptor-binding domain (RBD) and spike protein of the SARS-CoV-2 virus. Blue diamonds are measurements for negative samples, while red x are positive samples. The green rectangle is an example of confidence interval bounds often used to distinguish positive and negative samples. The yellow and blue domains correspond to* $\boldsymbol{D_P^\star}$ *and* $\boldsymbol{D_N^\star}$ *according to Eqs. (1) and (2). Contour lines are level sets of the conditional probability models P(r) and N(r). Note that the colored domains yield a more accurate classification relative to the green rectangle.*

From a scientific standpoint, the challenge of developing metrology tools for biological systems arises from the need to measure microscopic (cellular or sub-cellular) phenomena, compounded by the inherent variability and complexity of living organisms. For example, aspects of immune response are mediated via multiple types of lymphocytes (e.g., B-cells, T-cells, macrophages) and involve messaging and kinetic pathways that use molecules such as cytokines, growth factors, and antibodies [2]. Yet for characterizing previous exposure to a disease, assays often measure only one of these factors, i.e., antibodies. In some cases, however, cross-reactivity with immune response to other disease, inherent biological variability, and/or unexpected side-reactions can lead to false results; see, e.g., [3]. Such issues are further compounded by the fact that many biological concepts such as "immunity" are difficult, if not impossible to define quantitatively and objectively.

In many situations, these challenges can be reduced to ones of interpretation. The fundamental physical principles used to construct instrumentation such as photodetectors are well understood. Thus, one often assumes that the reported measurement values are accurate; e.g., the photodetector reported X photons because there were indeed that many. Instead, the relevant question is: does a value of X signify anything meaningful about the biological system? Answering this question requires inference about the behavior of a system from indirect observations. Thus, it fundamentally relies on tasks such as mathematical modeling and data analysis.

Optimal Classification for SARS-CoV-2 Antibody Testing. With the on-going COVID-19 health crisis, there is a growing need to identify individuals who have been infected by SARS-CoV-2 and to characterize immunity imparted by vaccines. In the medical community, it is well known that antibodies—'Y' shaped biomolecules that can attach to and disable viruses—play a primary role in immune response [2]. They also provide a lasting "memory" of how to fight a disease. Thus, antibody serology testing is often considered the gold standard for assessing previous infection. However, anecdotal and scientific evidence points to inaccuracy in SARS-CoV-2 assays being developed and suggests a need to revisit the underlying measurement science [3].

Key candidates for further investigation include mathematical classification schemes used to interpret assay results. In a typical serology test, the measurement apparatus records the amount of fluorescent light emitted by particles that attach to specific antibodies,

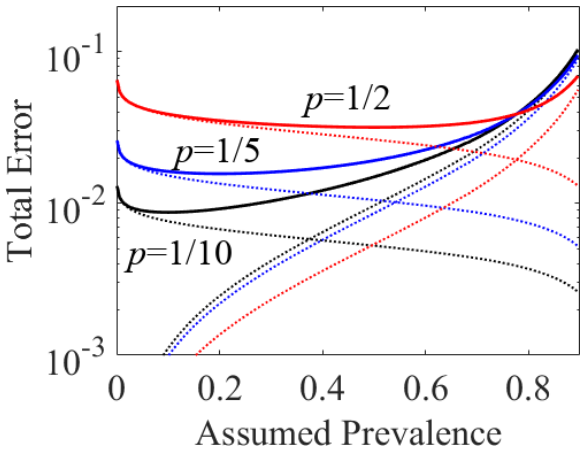


Figure 14. *Total error associated with using optimal domains* $\boldsymbol{D_P^\star}$ *and* $\boldsymbol{D_N^\star}$ *computed using an assumed prevalence q that differs from the true prevalence p. Note that the minimum error is achieved when the assumed and true prevalence values are equal.*

e.g., those having an affinity for the SARS-CoV-2 virus. However, the measurement value is only a number or a vector, which is clinically meaningless without further interpretation. Standard practice is to develop a *classification scheme* by using known positive and negative samples to identify ranges of measurement outcomes that correspond to these states. These classification schemes typically: (I) identify positive samples as those falling outside the confidence intervals for negative samples; or for scalar measurements (II) use a technique called receiver operating characteristics (ROC) to find a cutoff value (separating positives and negatives) that minimizes the false classification rate of the training data. See [3] and supporting documentation provided therein for examples.

ACMD staff members have developed a new classification framework that leverages optimal decision theory to improve testing accuracy. Critically, this work demonstrates that existing methods (listed above) increase error rates by up to a decade *due to sub-optimal data analysis alone*. The starting point of our analysis is the observation that the probability Q of a measurement outcome r can be decomposed using conditional probabilities via

$$Q(r) = pP(r) + (1-p)N(r)$$

where p is the disease prevalence, r is a scalar or vector associated with the raw measured value (e.g., in fluorescence units), $P(r)$ and $N(r)$ are the probabilities of a measurement outcome assuming the sample is positive or negative, and $Q(r)$ is the probability of any measurement outcome. The false classification rate can be expressed as

$$L[D_P, D_N] = (1-p)\int_{D_P} \mathrm{d}r\, N(r) + p\int_{D_N} \mathrm{d}r\, P(r)$$

where D_P *and* D_N are domains for which we classify samples as positive and negative, so that the terms on the right-hand side are the false positive and false negative rates. One can subsequently show (see Figure 13) that for smooth-enough conditional probabilities and up to sets of measure zero, this error rate is minimized by the domains

$$D_P^\star = \{r: pP(r) > (1-p)N(r)\} \quad (1)$$
$$D_N^\star = \{r: pP(r) < (1-p)N(r)\} \quad (2)$$

The domains $D_P^\star$ and $D_N^\star$ are notable for several reasons. For one, they generalize ROCs to multidimensional settings and arbitrary disease prevalence. This is interesting if only because ROC analysis is an 80-year-old technique originally developed to better detect aircraft during World War II; thus Eqs. (1) and (2) bring data analysis for serology testing up to modern standards. Just as importantly, the classification domains explicitly reveal the importance of disease prevalence, which is all but ignored in most assays; see Figure 14. These results also demonstrate that classification is fundamentally a task in mathematical modeling, since the optimal classification domains are automatically given once the conditional probabilities are known.

To further improve the usefulness of Eqs. (1) and (2), we have also developed an adaptive routine that yields unbiased estimates of disease prevalence in terms of the conditional probability models. Using serology data from a SARS-CoV-2 assay developed in [4], we have demonstrated that this algorithm can yield *optimal* classification to within negligible error using only a few thousand serology samples, which is well within the constraints of testing operations being established at citywide scales. Results of this work have been submitted for publication [5], and collaborations are on-going with the National Cancer Institute Serological Sciences Network program to transfer this technology to the serology community.

Data Analysis and Uncertainty Quantification for qPCR Testing. Quantitative PCR testing is among the most reliable means of detecting a disease at the early stages of infection [6]. To perform this measurement, a clinician first extracts from a patient a sample presumed to contain copies of the target pathogen, e.g., a virus. The qPCR instrument then cycles the sample through a sequence of 35 to 45 reactions, each time doubling the amount of DNA (which may be reversed transcribed from RNA) associated with the disease. Each new copy of DNA also adds a fluorescent molecule to the reaction mix. Thus, if the pathogen is present in the patient sample, exponential growth in the number of fluorophores will create a signal that can be detected by the qPCR instrument.

Despite this conceptually simple design, both qPCR assays and the instruments themselves suffer from effects that introduce systematic biases into fluorescence signals, which confounds attempts to

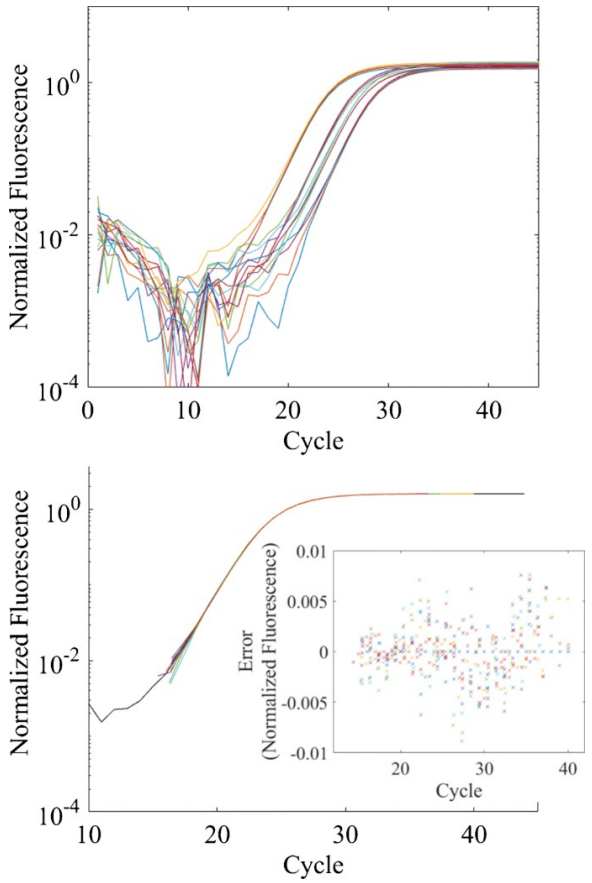


Figure 15. *qPCR amplification curves of the N1 fragment of a SARS-CoV-2 RNA construct. Top: qPCR curves after background subtraction. Bottom: curves after data collapse. The inset shows the error on an absolute scale relative to the master curve. The data collapse is accurate to better than 1% of the maximum scale of the data.*

analyze data. It is well known that background effects contribute 1 % or more to the signal and can be difficult to distinguish from late-amplification samples. However, the sources of this background have never been fully explored. Instead, a panoply of empirical methods has been developed in an attempt to subtract off such effects; see [7] and references contained therein. Nonetheless, anecdotal and published results point to the inadequacy of such approaches.

To address such issues, ACMD and MML staff members are collaborating with the US Geological Survey (USGS) to uncover and quantify the various contributions to background signals. The main idea behind this work is to design a hierarchy of experiments that sequentially add more components of the qPCR reaction mix and measure the resulting fluorescence profiles. Doing so yields a collection of "fingerprints" associated with the different background sources. We also find that the relative contributions of these sources vary according to their relative mole fractions, which has uncertainty due to pipetting errors. Thus, we have concurrently developed a constrained optimization framework that leverages the structure of previously characterized background signals to determine their relative contributions and subtract them off on the fly. Initial applications of these methods to both SARS-CoV-2 and freshwater testing of environmental DNA indicate up to a 10-fold improvement in measurement sensitivity of the fluorescence signals.

To realize this sensitivity improvement in terms of disease detection, we have also pursued data analysis and interpretation of qPCR curves viewed from the perspective of mathematical analysis. It is generally understood that the amount of DNA, $d_n(x, y)$, in a sample doubles with each cycle n, where x is the initial amount of DNA and y is a scalar denoting the concentration of other reactants, which are often in fixed proportion to one another. Thus, for constants $q > 0$, changing the initial amount of genetic material corresponds to transformations of the form

$$d_{n-q}(2^q, y) = d_n(1, y). \tag{3}$$

We concluded that the measurement process is also *linear* in the sense that the total measurement signal is the sum of signals arising from its parts. This implies a previously unknown property of qPCR signals, namely that they obey relationships of the form

$$d_n(k, ky) = k d_n(1, y) \tag{4}$$

where $k > 1$. Taken together, it is straightforward to show that Eqs. (3) and (4) imply that all qPCR curves of a fixed chemistry are identical up to an affine transformation of the form

$$d_n(\chi, \gamma) = a d_{n-b}(x, y) \tag{5}$$

where a and b are transformation parameters; see Figure 15. The result embodied in Eq. (5) is useful because it implies a method for identifying false positives. Conventional methods for analyzing qPCR amplification curves rely on the assumption that a true signal will cross a threshold. Given, however, that background signals often pollute the fluorescence signal, thresholds are often set to be significantly higher than the ultimate sensitivity of the photodetectors. In contrast to this, we use Eq. (5) to formulate an optimization framework in which we seek the transformation parameters that map a measurement signal onto a reference curve associated with pristine data. Constraints are used to determine whether this transformation can be achieved to within uncertainty of the instrumentation and thus reject those datasets that might otherwise be falsely classified as positive. Combining this approach with our background subtraction methods, we have demonstrated an ability to detect up to a 10-fold lower concentration of initial DNA. Moreover, we have demonstrated that this technique is applicable to the CDC qPCR assay for SARS-CoV-2. Results of this work have been published

in a manuscript in the journal *Analytical and Bioanalytical Chemistry* [7].

Future work aims to extend transfer this technology to the broader testing community and better characterize the physical processes that generate background signals.

References

[1] HR 266, Public Law 116-139, *Paycheck Protection Program and Health Care Enhancement Act*. April 24, 2020.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter. *Molecular Biology of the Cell 4th Edition.* New York, Garland Publishing, 2002.

[3] U.S. Food & Drug Administration. *EUA Authorized Serology Test Performance*. Web page[13], accessed 2020-09-16.

[4] T. Liu, J. Hsiung, S. Zhao, J. Kost, D. Sreedhar, C. V. Hanson, K. Olson, D. Keare, S. T. Chang, K. P. Bliden, P. A. Gurbel, U. S. Tantry, J. Roche, C. Press, J. Boggs, J. P. Rodriguez-Soto, J. G. Montoya, M. Tang, and H. Dai. Quantification of Antibody Avidities and Accurate Detection of SARS-CoV-2 Antibodies in Serum and Saliva on Plasmonic Substrates. *Nature Biomedical Engineering* **4** (2020), 1188.

[5] P. N. Patrone and A. J. Kearsley, Classification Under Uncertainty: Data Analysis for Diagnostic Antibody Testing, in review.

[6] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. W. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, D. G. J. C. Mulders, B. L. Haagmans, B. Van der Veer, S. Van den Brink, L. Wijsman, G. Goderski, J-L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. PG. Koopmans, C. Drosten. Detection of 2019 Novel Coronavirus (2019-ncov) by Real-time RT-PCR. Euro surveillance. *Bulletin Europeen sur les Maladies Transmissibles* 25:3 (2020), 2000045.

[7] P. N. Patrone, E. L. Romsos, M. H. Cleveland, P. M. Vallone, and A. J. Kearsley. Affine Analysis for Quantitative PCR Measurements. *Analytical and Bioanalytical Chemistry* **412** (2020), 7977–7988.

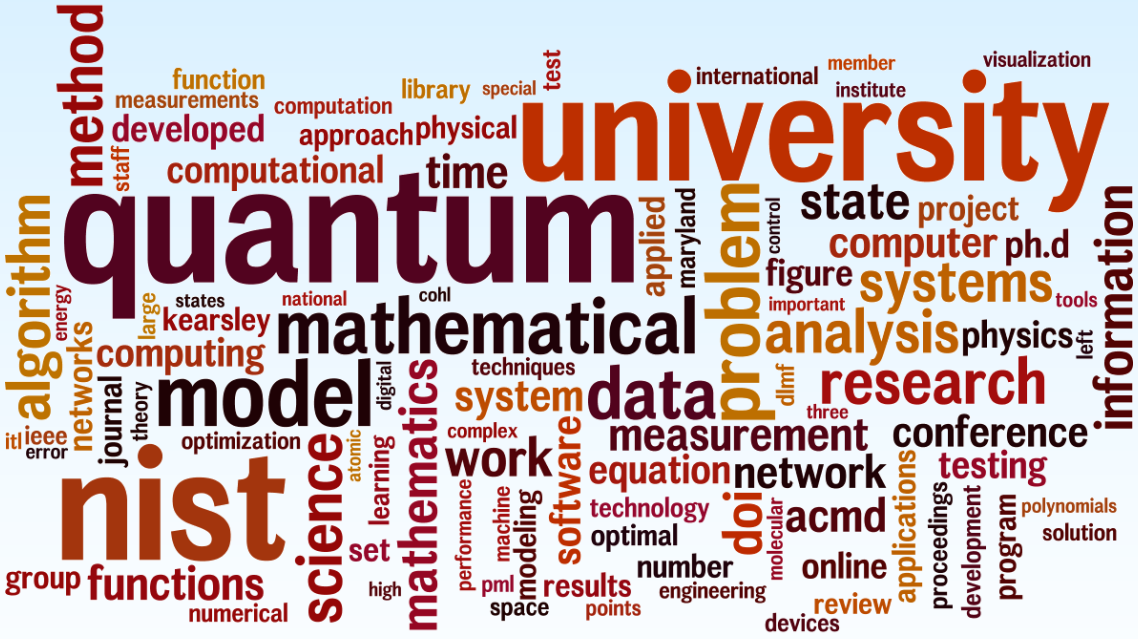
Participants

Anthony Kearsley, Paul Patrone, Robert DeJaco (ACMD); Peter Vallone, Erica Romsos (NIST MML)

[13] https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance

Part III

Project Summaries



Mathematics of Metrology

Mathematics plays an important role in measurement science. Mathematical models are needed to understand how to design effective measurement systems and to analyze the results they produce. Mathematical techniques are used to develop and analyze idealized models of physical phenomena to be measured, and mathematical algorithms are necessary to find optimal system parameters. Mathematical and statistical techniques are needed to transform measured data into useful information. We develop fundamental mathematical methods and tools necessary for NIST to remain a world-class metrology institute, and to apply these to measurement science problems.

Stabilized Richardson Leapfrog Scheme Run Backward in Time, and Explicit Stepwise Computation of Ill-Posed Time-Reversed 2D Navier-Stokes Equations

Alfred S. Carasso

Ill-posed deconvolution problems and associated time-reversed dissipative evolution equations pervade measurement science and are important in several other technological applications. In environmental forensics, much success has been achieved by solving advection diffusion equations backward in time to locate sources of groundwater contamination [1].

In numerous scientific measurements, the instrument point spread function is a bell-shaped distribution that may be well-approximated by a Gaussian, or by a heavy-tailed infinitely divisible probability density, often with parameters that are only tentatively known. Reformulating the integral equation deconvolution problem into an equivalent time-reversed generalized diffusion equation provides significant advantages. Marching backward in time stepwise, from a positive time T to time $t = 0$, allows the deconvolution to unfold in slow motion, provides the ability to monitor that process, and the possibility of terminating it prior to time $t = 0$ to prevent serious noise contamination and/or development of ringing artifacts. Such an approach, involving backward fractional and logarithmic diffusion equations, has been successfully applied in image deblurring of MRI and PET brain scans, nanoscale scanning electron micrographs, and galactic scale Hubble Space Telescope imagery [2-4].

More recently, these ideas have led to the development of a powerful new approach for solving ill-posed, time-reversed, multidimensional, nonlinear dissipative evolution equations, based on stabilizing explicit marching difference schemes. An appropriate easily synthesized compensating smoothing operator is applied at every time step to quench the instability. The stabilized scheme now becomes unconditionally stable, but slightly inconsistent, and eventually leads to a distortion away from the true solution. This is the stabilization penalty. However, in many problems of interest, the cumulative error is sufficiently small to allow for useful results. In a series of papers [5-11], such stabilized schemes were successfully applied to interesting classes of time-reversed nonlinear initial value problems for parabolic equations, viscous wave equations, coupled sound and heat flow, thermoelastic vibrations, 2D viscous Burgers equations, and most recently, 2D incompressible Navier-Stokes equations. Such computations had not previously been considered possible.

Leapfrog Scheme in Backward Navier-Stokes Equations. While quite good results are possible with the $O(\Delta t)$ explicit scheme discussed in [10], more accurate reconstructions may be feasible with an $O(\Delta t)^2$ explicit scheme, such as Richardson's leapfrog scheme. That scheme is notoriously unconditionally unstable in well-posed, forward, linear dissipative evolution equations. Remarkably however, the leapfrog scheme can be stabilized and marched backward in time, often providing more accurate results [11]. Effective smoothing operators based on $(-\Delta)^p$, with non-integer $p > 2$, can be efficiently synthesized using FFT algorithms. The analysis of numerical stability is restricted to a related linear problem. However, as is found in leapfrog computations of well-posed meteorological and oceanic wave propagation problems, such linear stability is necessary but not sufficient in the presence of nonlinearities. Here, likewise, additional Robert-Asselin-Williams (RAW) time-domain filtering [12], must be used to prevent characteristic leapfrog nonlinear instability, unrelated to ill-posedness.

The 2D Navier-Stokes problem is studied in *stream function-vorticity* formulation. In Figure 16, in the leftmost column, a Sydney Opera House image is used to create a stream function $\psi(x, y, 0)$ at $t = 0$, with contour plots shown in the middle leftmost image. The speeds $u = \psi_x$, $v = \psi_y$ and vorticity $\omega = -\Delta\psi$, are obtained by numerical differentiation, with contour plots of $\omega(x, y, 0)$ shown in the bottom leftmost image. Using a numerical method of modest accuracy, these initial data are evolved forward up to a time $T = 3 \times 10^{-3}$, producing the data shown in the middle column in Figure 16. With a kinematic viscosity $\nu = 0.01$, the

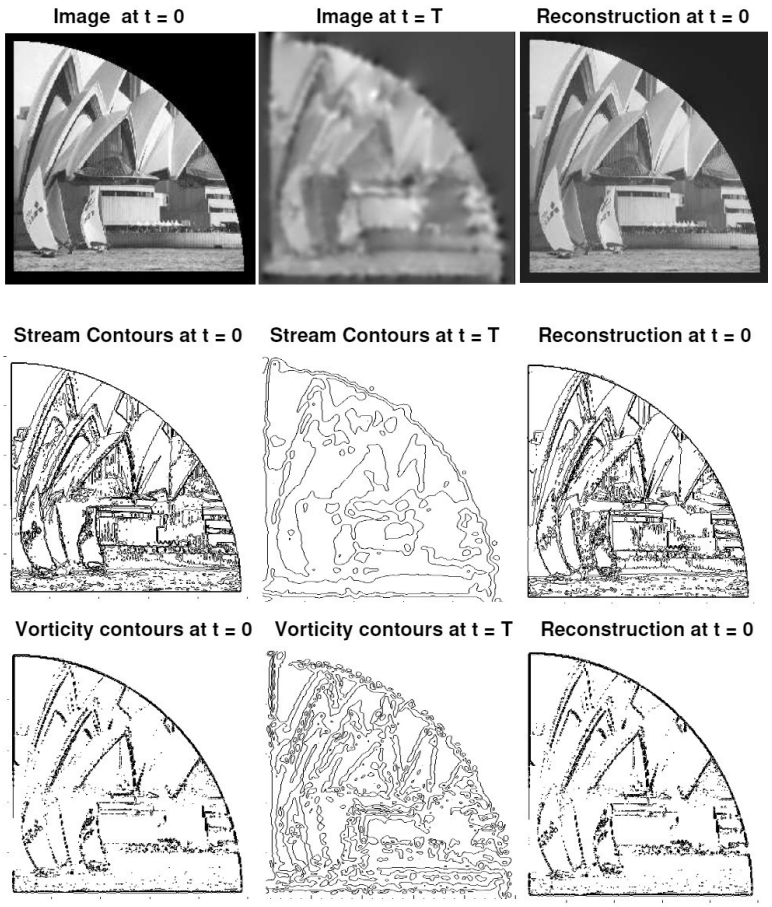


Figure 16. *Leapfrog backward recovery in Sydney data.*

associated Reynolds number $\mathrm{Re} = 11000$ in this experiment. Substantial erosion and disorganization of the edges and other sharp features in the leftmost column is now apparent in the middle column. The L_2 norms of u, v are reduced to 40 % of their initial values, while the L_2 norm of ω is reduced to 7 % of its initial value. Remarkably, as shown in the rightmost column of Figure 16, the stabilized leapfrog scheme, marching backward from the middle column data, can recover the data at $t = 0$ with good accuracy. The L_2 relative errors are 6.6 % for u, v, and 14.6 % for ω. Using the $O(\Delta t)$ scheme in [10] on the same middle column data, leads to slightly higher relative errors in u, v, but a 22 % relative error in ω. The value of T in this successful experiment is significantly higher than would seem feasible on the basis of the best-known uncertainty estimates for backward Navier-Stokes solutions [13].

While the leapfrog scheme often provides more accurate vorticity reconstructions than the scheme in [10], examples have been found where the opposite is true. Thus, both schemes are valuable computational tools, and both should be tried in any problem where either scheme is found useful.

[1] J. Atmadja and A. C. Bagtzoglou. State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification. *Environmental Forensics* **2** (2001), 205-214.

[2] A. S. Carasso. The APEX Method in Image Sharpening and the Use of Low Exponent Lévy Stable Laws. *SIAM Journal on Applied Mathematics* **63** (2002), 593-618.

[3] A. S. Carasso, D. S. Bright, and A. E. Vladar. APEX Method and Real-Time Blind Deconvolution of Scanning Electron Microscope Imagery. *Optical Engineering* **41** (2002), 2499-2514.

[4] A. S. Carasso. Bochner Subordination, Logarithmic Diffusion Equations, and Blind Deconvolution of Hubble Space Telescope Imagery and Other Scientific Data. *SIAM Journal on Imaging Sciences* **3** (2010), 954-980.

[5] A. S. Carasso. Stable Explicit Time-Marching in Well- Posed or Ill-Posed Nonlinear Parabolic Equations. *Inverse Problems in Science and Engineering* **24** (2016), 1364-1384.

[6] A. S. Carasso. Stable Explicit Marching Scheme in Ill- Posed Time-Reversed Viscous Wave Equations. *Inverse Problems in Science and Engineering* **24** (2016), 1454-1474.

[7] A. S. Carasso. Stabilized Richardson Leapfrog Scheme in Explicit Stepwise Computation of Forward or Backward Nonlinear Parabolic Equations. *Inverse Problems in Science and Engineering* **25** (2017), 1719-1742.

[8] A. S. Carasso. Stabilized Backward in Time Explicit Marching Schemes in the Numerical Computation of Ill-Posed Time-Reversed Hyperbolic/Parabolic Systems. *Inverse Problems in Science and Engineering* **27** (2019), 134-165.

[9] A. S. Carasso. Stable Explicit Stepwise Marching Scheme in Ill-Posed Time-Reversed 2D Burger's Equation. *Inverse Problems in Science and Engineering* **27** (2019), 1672-1688.

[10] A. S. Carasso. Computing Ill-Posed Time-Reversed 2D Navier-Stokes Equations, using a Stabilized Explicit Finite Difference Scheme Marching Backward in Time. *Inverse Problems in Science and Engineering* **28**:7 (2019), 988-1010.

[11] A. S. Carasso. Stabilized Leapfrog Scheme Run Backward in Time, and the Explicit $O(\Delta t)^2$ Stepwise Computation of Ill-Posed Time-Reversed 2D Navier-Stokes Equations. In Review.

[12] P. D. Williams. The RAW Filter: An Improvement to the Robert-Asselin Filter in Semi-Implicit Integrations. *Monthly Weather Review* **139** (2011), 1996-2007.

[13] R. J. Knops and L. E. Payne. On the Stability of Solutions of the Navier-Stokes Equations Backward in Time. *Archives of Rational Mechanics and Analysis* **29** (1968), 331-335.

Numerical Solutions of the Time Dependent Schrödinger Equation

Barry I. Schneider
Heman Gharibnejad (Computational Physics Inc.)
Luca Argenti (University of Central Florida)
Juan Martin Randazzo (University of Central Florida)
Nicholas Douguet (Kennesaw State University)
Jeppe Olsen (Aarhus University)
Ryan Schneider (University of California San Diego)

We have been collaborating with various scientists for several years developing numerically robust methods for solving the time-dependent Schrödinger equation (TDSE). Luca Argenti, Nico Douguet, and Juan Martin Randazzo joined the effort in 2019. Heman Gharibnejad, a former NIST NRC postdoctoral associate, has also continued to work on the project part time. There are three related research threads underway.

- Developing a hybrid finite element discrete variable (FEDVR)-Gaussian approach to treat the interaction of attosecond (10^{-18} sec) radiation with molecular targets.
- Computing the required hybrid matrix elements via a novel 3D numerical grid.
- Examining the performance of various numerical time propagation techniques for the TDSE.

The hybrid finite element approach that we have developed is quite general, but the applications to date have concentrated on describing the single and double ionization of electrons exposed to intense, ultrafast, laser radiation in many-body atomic and molecular systems. These attosecond (10^{-18} sec) pulses provide a new window to study the electronic motion in atoms and molecules on their natural timescale. To put this in context, the motion of electrons responsible for chemical binding and electron transfer processes in nature have a characteristic timescale of about 100 attoseconds. (It takes an electron 152 attoseconds to go around the hydrogen atom.) These processes can only be described using time-dependent quantum mechanics. Where appropriate, this needs to be coupled to Maxwell's equations to describe macroscopic phenomena. Our overall goal is to image quantum phenomena with sub-femtosecond temporal and sub-Angstrom spatial resolution and to provide coherent control of electron dynamics. Eventually, one can contemplate producing "molecular movies" of this motion in much the same way as it is done in molecular dynamics simulations of heavy particle processes.

The basic methodology as applied to atoms and simple diatomic molecules has been described in [1-11]. Reference [4] provides a detailed review of the work. The essential aspects have been

- development of the finite element discrete variable method (FEDVR) to spatially discretize the coordinates of the electrons,
- construction of numerical grid capable of efficiently computing the required one and two electron matrix elements, and
- extensions of the short iterative Lanczos method to propagate the wavefunction in time.

We have efficiently parallelized the FEDVR method using MPI, shown that it scales linearly with the size of the FEDVR basis and applied it to selected problems [1-13]. Large-scale calculations have been performed on a number of atoms and molecules using resources provided by the NSF Extreme Science and Engineering Discovery Environment (XSEDE) program.

More recently we have begun to employ a mixed basis of Gaussian functions at short range and FEDVR functions at long range to extend our methods to complex polyatomic molecules. This approach has several important advantages over using a single basis over all of space. First, the use of nuclear-centered Gaussians preserves the local atomic symmetry around each nucleus and avoids the poor and often non-convergent behavior of using a single-center FEDVR basis at all distances. Second, once the electron is far enough away from the nuclear cusps, a single center expansion converges quickly and, importantly, can represent the electrons out to very large distances using an approach that is very amenable to domain decomposition.

The major issue is to compute the one and two electron integrals between the two types of basis functions. The formalism we have been developing requires as input, transition density matrices extracted from a high-level quantum chemistry code, in order to compute the additional one and two-electron integrals. J. Olsen, who has experience in such computations, has been collaborating with us on the project. The NIST-UCF-KSU-Aarhus group meets on a weekly basis to discuss the issues and plans. During the past year, the density matrices have been integrated into a new code called ASTRA and some preliminary tests of the code have been performed that indicate things are working as expected. It should be noted that this is a very complex many-body problem and even with the most talented of researchers, it is a long-term effort.

The calculation of the one and two-electron matrix elements over the hybrid basis must be performed numerically. Given the polycentric nature of the electron distribution and the need to compute these integrals to significantly larger distances than in quantum chemistry calculations, it was necessary to develop an efficient 3D integration scheme. One cannot use methods which fix the coordinate system at a single point in space, as they are at best very slowly convergent and often do not converge at all. To overcome that, a popular approach developed by Axel Becke, based on the partition of

unity, defines atomic grids centered on the atoms where the grid points are appropriately weighted to satisfy the partition of unity. The original method of Becke required substantial modification for our purposes. We require a central grid that can describe the much larger integration region without too much contamination from the points on the nuclear centers. To accomplish that, we define a new partition of unity that constrains the atomic grid points to small atomic spheres. A central grid is then added to take care of the interstitial and longer-range parts of configuration space where the atomic grid points are forced to vanish. The new partition of unity does remarkably well in performing very accurate integration for integrands having nuclear cusps as well as oscillating at larger distances. A paper describing the approach is in review [14].

Lastly, we have been engaged in efforts to generalize the short iterative Lanczos (SIL) method used to integrate the TDSE to efficiently and accurately integrate for much longer times [15]. An integral equation formalism has been developed which exploits the fact that part of the Hamiltonian is a linear operator and can be treated using an exponential propagator, which is exactly what the SIL provides. The remainder term, involving a time integral over the residual interaction and the unknown wavefunction, is computed by numerical integration over the large time step using the SIL on each of the terms in the integrand. To do the integral requires that we have a previous approximation to the wavefunction. Thus, the entire scheme is iterative starting with the wavefunction from the previous time step.

This summer, we were lucky enough to enlist a mathematics graduate student, Ryan Schneider[14], to work on the problem. Ryan was supported by the NSF Mathematical Sciences Graduate Internship (MSGI) program and worked remotely with us for the summer. We have continued to work with Ryan since the end of the internship as the work was not finished and Ryan was eager to continue working on the problem part time.

For completeness, we also reference a review paper [16] by Schneider and Gharibnejad which received the ITL Outstanding Journal Paper award in 2020.

[1] J. Feist, S. Nagele, R. Pazourek, E. Persson, B. I. Schneider, L. A. Collins, and J. Burgdörfer. Nonsequential Two-Photon Double Ionization of Helium. *Physical Review A* **77** (2008), 043420.

[2] X. Guan, K. Bartschat, and B. I. Schneider. Dynamics of Two-photon Ionization of Helium in Short Intense XUV Laser Pulses. *Physical Review A* **77** (2008), 043421.

[3] X. Guan, K. Bartschat, and B. I. Schneider. Two-photon Double Ionization of H_2 in Intense Femtosecond Laser Pulses. *Physical Review A* **82** (2010), 041407.

[4] B. I. Schneider, J. Feist, S. Nagele, R. Pazourek, S. Hu, L. Collins, and J. Burgdörfer. Recent Advances in Computational Methods for the Solution of the Time-Dependent Schrödinger Equation for the Interaction of Short, Intense Radiation with One and Two Electron Systems, in Dynamic Imaging. In *Quantum Dynamic Imaging* (A. Bandrauk and M. Ivanov eds.), CRM Series in Mathematical Physics, Springer, New York, 2011.

[5] X. Guan, E. Secor, K. Bartschat, and B. I. Schneider. Double-slit Interference Effect in Electron Emission from H_2^+ Exposed to X-Ray Radiation. *Physical Review A* **85** (2012), 043419.

[6] X. Guan, K. Bartschat, B. I. Schneider, L. Koesterke, Resonance Effects in Two-Photon Double Ionization of H_2 by Femtosecond XUV Laser Pulses. *Physical Review A* **88** (2013), 043402.

[7] J. Feist, O. Zatsarinny, S. Nagele, R. Pazourek, J. Burgdörfer, X. Guan, K. Bartschat, and B. I. Schneider. Time Delays for Attosecond Streaking in Photoionization of Neon. *Physical Review A* **89** (2014), 033417.

[8] X. Guan, K. Bartschat, B. I. Schneider, and L. Koesterke. Alignment and Pulse-duration Effects in Two-photon Double Ionization of H_2 by Femtosecond XUV Laser Pulses. *Physical Review A* **90** (2014), 043416.

[9] B. I. Schneider, L. A. Collins, X. Guan, K. Bartschat, and D. Feder. Time-Dependent Computational Methods for Matter Under Extreme Conditions. *Advances in Chemical Physics* **157** (2015), Proceedings of the 240 Conference: Science's Great Challenges, (A. Dinner, ed.), John Wiley.

[10] B. I. Schneider, X. Guan, and K. Bartschat. Time Propagation of Partial Differential Equations Using the Short Iterative Lanczos Method and Finite-Element Discrete Variable Representation. *Advances in Quantum Chemistry* **72** (2016), 95-127.

[11] B. I. Schneider. How Novel Algorithms and Access to High Performance Computing Platforms are Enabling Scientific Progress in Atomic and Molecular Physics. *Journal of Physics: Conference Series* **759** (2016), 012002.

[12] B. I. Schneider. 45 Years of Computational Atomic and Molecular Physics: What Have We (I) Learned. *Journal of Physics: Conference Series* **875** (2017).

[13] B. I. Schneider, L. A. Collins, K. Bartschat, X. Guan, and S. X. Hu. A Few Selected Contributions to Electron and Photon Collisions with H_2 and H_2^+. *Journal of Physics* **50** (2017), 214004.

[14] H. Gharibnejad, N. Douguet, B. I. Schneider, J. Olsen and L. Argenti, A Multi-Center Quadrature Scheme for the Molecular Continuum. In review.

[15] H. Gharibnejad, B. I. Schneider, M. Leadingham, and H. J. Schmale. A Comparison of Numerical Approaches to the Solution of the Time-Dependent Schrödinger Equation in One Dimension. *Computer Physics Communications* (2019), 106808, to appear.

[16] B. I. Schneider and H. Gharibnejad. Numerical Methods Every Atomic and Molecular Theorist Should Know. *Nature Reviews Physics* **2** (2020), 89-102.

[14] No relation to the project PI.

A Science Gateway for Atomic and Molecular Physics

Barry I. Schneider
Lincoln Carr (Colorado School of Mines)
Klaus Bartschat (Drake University)
Kathryn Hamilton (Drake University)
Oleg Zatsarinny (Drake University)
Igor Bray (Curtin University, Australia)
Armin Scrinzi (Ludwig-Maximilians U., Germany)
Fernando Martıın (U. Autonoma de Madrid, Spain)
Jesus Gonzalez Vasquez (U. Autonoma de Madrid)
Jimena Gorfinkiel (Open University, UK)
Robert Lucchese (Lawrence Berkeley Laboratory)
Sudhakar Pamidighantam (Indiana University)
Nicholas Douguet (Kennesaw State University)
Samantha Fonseca dos Santos(Rollins College)

An international effort has been underway since 2018 to develop and maintain a Science Gateway for Atomic and Molecular Physics (AMP) [1, 2]. The purposes of the gateway are to

- collect and make available to the community a set of advanced computational tools that are actively being used to study atomic and molecular collisions and the interaction of radiation with atoms and molecules,
- provide educational materials for beginning and advanced users desiring to learn the ideas and concepts of AMP, both theoretical and computational and
- make available to the broad community atomic and molecular data needed for many applications. The availability of collision data is critical to many areas of physics including astrophysics, fusion energy, the study of lighting and microelectronics.

Codes for modeling and simulation of such phenomena have been developed in specific groups by graduate students and postdocs but are often poorly documented, and unavailable outside the group developing the software. This leads to "reinventing the wheel" in too many instances. Maintaining these computational tools, as well as enhancing their capabilities, is one of the major goals of the project and is critical to ensure continued rapid scientific progress in AMP.

Another important goal is to enable the code developers themselves to compare the calculations of specific well-defined problems using different methodologies. This enables the verification of results of different codes and encourages comparison with experiment, when available. It has already been demonstrated that a few of these codes are often more accurate than experiment and thus provide a predictive capability when experimental results are unavailable.

At the outset, the group acknowledged that, in contrast to some other communities, the AMP community has lagged behind in developing community supported software packages that are robust and used by others. The group was convinced the time had arrived to change existing practices and make these tools available and easily used by future generations of AMP scientists as well as the developers themselves.

The group wrote a proposal to the NSF XSEDE program to fund some initial development of the gateway. The proposal was successful and, importantly, provided the developers with some hands-on assistance from the Extended Collaborative Support Services arm of XSEDE. This was vital to the success of the effort. In particular, we acknowledge the important contribution of Sudhakar Pamidighantam of IU in making our efforts a success.

An AMP Gateway workshop, the second such workshop, was held at NIST on December 11-13, 2019. The workshop was jointly supported by the Molecular Software Science Institute (MOLSSI) at Virginia Tech, the NSF, and NIST. Some 30 participants came together to discuss what has been accomplished and to present ideas for the future. The developers gave presentations on the science behind the codes and how to use the codes via the gateway. Pamidighantam spoke on the details of the gateway and how it can be used effectively, even for individuals desiring results but uninterested in the specifics of a given code. Other participants presented details of codes that have not yet been implemented on the gateway, and several have now expressed strong interest in joining the project. The workshop concluded with a discussion of what was needed for the future.

There have been some important advances during the current reporting period. All of the six major codes chosen for initial deployment have been ported to at least three XSEDE supercomputers and a new code, R-matrix with time, has been added to the portfolio. There is an early version of an API available to enable users to perform calculations with these codes. While the current API interface is still in its infancy, we are aware of what needs to be done and progress is ongoing. The need for longer term support for the project is critical to its ultimate success. To that end, the group spent considerable effort during the fall of 2020 to write an NSF proposal to the Cyberinfrastructure for Sustained Innovation program. L. Carr, of the Colorado School of Mines, joined the project and agreed to serve as PI. There are Co-PI's from IU (S. Pamidighantam) and Rollins College (S. Fonseca dos Santos). In addition, we recruited N. Douguet, from Kennesaw State, who joined Schneider and Bartschat, as Senior Investigators. The proposal was submitted from the Colorado School of Mines as the lead organization. The involvement of the Colorado School of Mines in the project has enabled us to get needed resources to develop the portal pages into what

is rapidly becoming a very professional state. The current portal contains good descriptions of the codes, the people involved, links to documentation, a bibliography, a preliminary data repository and some graphical material illustrating a few of the calculations that have been done with the codes.

[1] B. I. Schneider, K. Bartschat, O. Zatarinny I. Bray, A. Scrinzi, F. Martin, M. Klinker, J. Tennyson, J. Gorfinkiel, and S. Pamidighanta. A Science Gateway for Atomic and Molecular Physics. Preprint arXiv:2001.02286. URL: https://arxiv.org/abs/2001.02286

[2] B. I. Schneider, K. Bartschat, O. Zatsarinny, K. Hamilton, I. Bray, A. Scrinzi, F. Martin, J. G. Vasquez, J. Tennyson, J. Gorfinkiel, R. Lucchese and S. Pamidighantam. Atomic and Molecular Scattering Applications in an Apache Airavata Science Gateway. In *Proceedings of PEARC '20: Practice and Experience in Advanced Research Computing*, Online, July 2020. DOI: 10.1145/3311790.3397342

Regularization of a Smooth Function Defined by Noisy Data

Timothy Burns

Assume that $g = g(x)$ is a smooth data function on $[a, b]$, but we are only given a discrete noisy sample of the function, with unknown but uniform 2σ error bars, at equally spaced data points,

$$\mathbf{g}^{\epsilon} = \mathbf{g} + \epsilon = \{g_1^{\epsilon}, \ldots, g_m^{\epsilon}\},\ a < x_1 < \cdots < x_m = b.$$

There are a number of methods that can be applied to regularize the data function, such as nearest-neighbor averaging, which can be interpreted as smoothing by means of a low-pass filter that is connected with a diffusion process, which preferentially dampens higher-frequency noise in the data.

An alternative regularization method that is being studied here originated as a collaboration with former NIST colleague B. Rust [1]. The key idea is to treat the separation of signal from noise in $g(x)$ as an inverse problem, which includes obtaining an estimate of the standard deviation σ as part of the solution. The procedure is to formulate the inverse problem as a Volterra integral equation of the first kind.

$$Af = \int_a^x f(y)\, dy = g(x), \qquad g(a) = 0.$$

In this setting, we want to obtain an estimate of the *source* function $f = f(x)$, which amounts to finding an approximation of the derivative of g,

$$A^{-1}g = \frac{dg}{dx} = f(x),$$

under the assumptions that the residual corresponding to the noise in the data is a realization of a white noise time series, and the measurements are from a process which preferentially damps high-frequency content in the data. By means of a singular value decomposition of the integration operator [1, 2, 3], we use a finite-dimensional spectral projection method that determines a smooth, closed-form regularization of the data function g, and a corresponding closed-form estimate of the source function f, the derivative of g. The standard deviation σ is determined by selecting the value for which the noise is closest to a normal distribution.

[1] T. J. Burns and B. W. Rust. Closed-Form Projection Method for Regularizing a Function Defined by a Discrete Set of Noisy Data and for Estimating its Derivative and Fractional Derivative. arXiv preprint 1805.09849v1.

[2] Z. Zhao, Z. Meng, and G. He. A New Approach to Numerical Differentiation, *Journal of Computational and Applied Mathematics* **232** (2009), 227-239.

[3] R. Gorenflo and K. V. Tuan. Singular Value Decomposition of Fractional Integration Operators in L_2-Spaces with Weights. *Journal of Inverse and Ill-Posed Problems* **3** (1995), 1-10.

Seminorm Regularization of Linear Inverse Problems

Matthew J. Roberts
Mark S. Gockenbach (University of Delaware)

Linear inverse problems arise in many applications pertaining to measurement. In general, a linear inverse problem is posed as follows: Determine solution x in the equation

$$T\,x = y,$$

where y is a measurement of y^*. Here, x, y, and y^* are usually functions in a function space such as $L^2(0, 1)$, and this is the setting we will assume. The basic question is the following: if y is a good estimate of y^*, then is x a good estimate of the true solution x^* ? Here,

$$T\,x^* = y^*.$$

The inverse problems of interest arise when the solution x does not continuously depend on the data y. This means that an accurate measurement y of y^* can still result in a large distance between x and x^*. An inverse problem of this type occurs in the case that T is an integral operator over a closed and bounded domain, and more generally, whenever T is a compact operator (see [2] and [5]). One way to deal with this is by using seminorm regularization. We consider the following problem: Determine solution $\tilde{x}$ to the problem

$$\min_{u \in L^2(0,1)} |Tu - y|_{L^2(0,1)^2} + \lambda |Lu|_{L^2(0,1)^2}. \qquad (1)$$

Here, L is an operator (such as the derivative operator), and λ is a scalar greater than 0. The solution $\tilde{x}$ is the seminorm regularized solution. The idea is to choose an operator L that reflects an undesirable feature in the solution x. For example, if the solution to (1) is highly oscillatory, a good choice for L would be the derivative operator.

The generalized singular value expansion (GSVE) can be used to simultaneously diagonalize the operator pair (T, L) in problem (1) (see [3], sections 2 through 4). This makes determining the solution $\tilde{x}$ of (1) transparent. In practice the operator pair (T, L) is discretized by the pair of finite dimensional operators (T_h, L_h), where, as $h \to 0$, the operator pair (T_h, L_h) goes to (T, L) in some sense. It was shown in [4] that the GSVE of (T_h, L_h) converges to the GSVE of (T, L) as $h \to 0$ if $T_h \to T$ and $L_h \to L$ in the operator norm as $h \to 0$.

Building on this work, together with results from [1], we have been able to derive the rates of convergence of the approximate generalized singular vectors of (T_h, L_h) to the true generalized singular vectors of (T, L). Numerical examples show that these derived rates of convergence are the best possible in the case of variational approximation. Such results will be most useful in error analysis pertaining to seminorm regularization. We currently have a manuscript with these results that is very close to complete.

[1] D. K. Crane, M. S. Gockenbach, and M. J. Roberts Approximating the Singular Value Expansion of a Compact Operator. In review.

[2] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publishers Group, Dordrecht, 1996.

[3] M. S. Gockenbach. Generalizing the GSVD. *SIAM Journal on Numerical Analysis* **54** (2016), 2417- 2540.

[4] M. S. Gockenbach and M. J. Roberts. Approximating the Generalized Singular Value Expansion. *SIAM Journal on Numerical Analysis* **56**:5 (2018), 2776-2795.

[5] M. S. Gockenbach. *Linear Inverse Problems and Tikhonov Regularization,* MAA Press, 2016.

Computational Tools for Image and Shape Analysis

Günay Doğan
Javier Bernal
James Lawrence
Charles R. Hagwood (NIST ITL)
Prashant Athavale (Clarkson University)
Harbir Antil (George Mason University)
Soeren Bartels (University of Freiburg)
Marilyn Y. Vazquez (Ohio State University)
Shuang Li (University of Southern California)
Eve N. Fleisig (Princeton University)
Olakunle Abawonse (SUNY Binghamton)

The main goal of this project is to develop efficient and reliable computational tools to detect geometric structures, such as curves, regions and boundaries from given direct and indirect measurements, e.g., microscope images or tomographic measurements, as well as to evaluate and compare these geometric structures or shapes in a quantitative manner. This is important in many areas of science and engineering, where the practitioners obtain their data as images, and would like to detect and analyze the objects in the data. Examples include microscopy images for cell biology, micro-CT (computed tomography) images of microstructures in material science, and shoeprint images in crime scenes for footwear forensics.

In fiscal year 2020, advances were made in both fronts of this project: image segmentation and shape analysis. Python implementations of solutions for problems in these areas were released, and documentation and examples were provided. A Matlab code for shape analysis was also released. In the following, we provide more details on the specific work carried out.

Image Segmentation. Image segmentation is the problem of finding distinct regions and their boundaries in given images. It is a necessary data analysis step for many problems in cell biology, forensics, and material science, as well as other fields in science and engineering. In FY 2020, G. Doğan and his collaborators continued to work on multiple strategies for image segmentation.

The first of these strategies builds on implicit representations of image regions by embedding them in phase field functions, developed previously by Doğan, and his collaborators, H. Antil of George Mason University and S. Bartels from the University of Freiburg [1]. Initial guesses of region candidates are placed on the image and the corresponding phase field function is evolved in a manner optimizing a data fidelity and boundary regularity. This algorithm tends to produce well-defined segmentations without stray pixels. Bug fixes and improvements to the numerical algorithm were introduced,

and Doğan started implementing a 3d version of the algorithm to segment volumetric images.

Doğan, and former guest researcher S. Li of USC had previously developed a novel image segmentation algorithm using topology optimization. This algorithm, which is suitable for multiphase image segmentation, performs image segmentation by iteratively assigning region labels to image pixels. It distinguishes regions by assuming distinct parametric distributions of pixel values within regions. It incorporates a nonlocal geometric regularization term in the optimization cost function to achieve well-defined regions. Doğan and Li performed further experiments and benchmarks to demonstrate the effectiveness of this approach, and wrote a paper describing the algorithm and its performance [3].

Doğan worked with graduate summer intern Olakunle Abawonse of SUNY Binghamton to implement an image segmentation algorithm that leverages a convexified version of a piecewise constant segmentation energy. They were able to achieve successful two-phase segmentations with this approach. They are currently working on a multiphase extension of this algorithm. Doğan also worked with former NIST guest researcher Marilyn Vazquez of Ohio State University to finish a manuscript on application of manifold learning algorithm to image segmentation [2].

Doğan continued his collaboration with Prashant Athavale of Clarkson University on preprocessing and segmentation of orientation images of microstructures. Orientation images, more specifically electron backscatter diffraction (EBSD) images, often come with many misorientation pixels, which have the appearance of noise, and may have regions of missing data. To alleviate these issues, they have been developing PDE-based denoising and inpainting algorithms to produce high quality reconstructions of the EBSD images. They compared their PDE-based algorithms with other denoising, and data reconstruction algorithms commonly used for EBSD images, and have been able to produce state-of-the-art reconstructions.

Shape Analysis. A natural approach to analyze and compare objects in image and data is through their shapes, an aspect that is invariant to rotation, translation and scaling. Shape-based analysis can facilitate object recognition and can provide a more intrinsic way to perform statistics of geometric data. Previously, Doğan, along with J. Bernal and C. Hagwood, developed an efficient optimization algorithm to compute the elastic shape distances between 2d closed curves [4, 5, 6, 7]. This shape distance is the fundamental building block for shape analysis.

Research and development for the computation of the elastic registration of two simple curves in *d*-dimensional space (and thus the associated elastic shape distance between them) was successfully completed with Bernal's lead [8, 9, 10]. For details, see the next project report below.

The investigation of a related problem is underway, that of computing the elastic registration of three-dimensional parametrized surfaces, to identify a "shape space" of surfaces together with a numerical framework for the computation of geodesics in such a space. Parallel to these developments, Doğan worked on a Python implementation of the elastic shape distance algorithm, developed in FY 2019 with K. Su, a student intern from Stanford University. He tested and benchmarked the code to release as part of the scikit-shape package.[15]

Using different shape representations or different versions of the algorithms lead to different shape dissimilarity metrics, and this brings the question of which metric would perform best. Doğan and student intern E. Fleisig of Princeton University had been developing a Python program, VEMOS (Visual Explorer for Metrics Of Similarity) that can be used to evaluate and compare multiple competing similarity/dissimilarity metrics, including shape dissimilarity metrics. VEMOS is useful for applications beyond shape distances; it can be used in a versatile manner to evaluate multiple alternative dissimilarity metrics for heterogeneous data sets, including images, shapes, point clouds and other data types. In FY 2020, Doğan continued to develop VEMOS, made bug fixes, and demonstrated it on example data sets for shoeprint forensics. They completed a manuscript describing VEMOS and submitted to a journal for review [11].

[1] H. Antil, S. Bartels, and G. Doğan. A Phase Field Segmentation Model with Improved Boundary Regularization. In preparation.

[2] M. Vazquez, T. Sauer, T. Berry, and G. Doğan. Texture Segmentation from a Manifold Learning Perspective. In preparation.

[3] S. Li and G. Doğan. Image Segmentation by Topology Optimization of Region Statistics. In review.

[4] A. Srivastava, E. Klassen, S. Joshi, and I. Jermyn. Shape Analysis of Elastic Curves in Euclidean Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**:7 (2011), 1415-1428.

[5] G. Doğan, J. Bernal, and C. R. Hagwood. Fast algorithms for Shape Analysis of Planar Objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitio*n (CVPR'15), Boston, MA, June 2015.

[6] G. Doğan, J. Bernal, and C.R. Hagwood. *FFT-based* Alignment of 2d Closed Curves with Application to Elastic Shape Analysis. In *Proceedings of the 1st International Workshop on Differential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories* (DiffCV'15), Swansea, United Kingdom, September 2015.

[15] https://github.com/scikit-shape/scikit-shape, http://scikit-shape.org/

[7] J. Bernal, G. Doğan, and C.R. Hagwood. Fast Dynamic Programming for Elastic Registration of Curves. In *Proceedings of the 2nd International Workshop on Differential Geometry in Computer Vision and Machine Learning* (DiffCVML'16), Las Vegas, NV, July 1, 2016.

[8] J. Bernal, J. Lawrence, G. Doğan, and C.R. Hagwood. On Computing Elastic Shape Distances Between Curves in *d*-Dimensional Space. In review.

[9] J. Lawrence, J. Bernal, and C. Witzgall. A Purely Algebraic Justification of the Kabsch-Umeyama Algorithm. *Journal of Research of the National Institute of Standards and Technology*, **124** (2019), 124028.

[10] J. Bernal and J. Lawrence. Characterization and Computation of Matrices of Maximal Trace Over Rotations. *Journal of Geometry and Symmetry in Physics* **53** (2019), 21-53

[11] E. Fleisig and G. Doğan. VEMOS: GUI for Evaluation of Distance Metrics of Heterogeneous Data Sets. In review.

Computing Elastic Shape Distances Between Curves in Higher Dimensions

Javier Bernal
James Lawrence
Günay Doğan
Charles Hagwood (NIST ITL)

A key problem in shape analysis [3, 10] is the computation of the elastic registration of two simple curves in *d*-dimensional space, where *d* a positive integer, and thus the associated elastic shape distance between them. We recently completed work on this problem; a manuscript has been prepared for publication [1] and a software package, associated with this work has been made available.[16] In our work we have enhanced Srivastava et al.'s previous contributions [10, 11] in various ways.

First a dynamic programming algorithm has been developed and implemented that is linear for computing an optimal diffeomorphism for the elastic registration of two simple curves in *d*-dimensional space. The computation of the registration is based only on reparameterizations (with diffeomorphisms of the unit interval) of one of the curves (no rotations). The curves are given on input as discrete sets of nodes in the curves, the numbers of nodes in the curves not necessarily equal, and the partitions of the unit interval discretizing the curves not necessarily uniform. Note this algorithm was already presented in [2] for d = 2.

Next, a purely algebraic justification has been developed of the usual algorithm, the Kabsch-Umeyama algorithm [7, 8, 12], for computing an optimal rotation matrix for the rigid alignment of two simple curves in *d*-dimensional space. Here the curves are given on input as discrete sets of nodes in the curves, the same number of nodes in each curve, the two curves discretized by the same partition of the unit interval, and the partition discretizing the curves not necessarily uniform. Note that computing this optimal rotation matrix is an instance of solving the so-called Wahba's problem [4, 13], especially if the partition discretizing the curves is not uniform. Note as well that the algebraic justification of the Kabsch-Umeyama algorithm was already presented in [9] but without the development of the matrix on which the algorithm is applied when a nonuniform partition is involved.

Lastly, with the convention that if one of the curves is closed, the first curve is closed, the L^2 we have redefined the distance that is minimized in Srivastava et al.'s work to allow for the second curve to be reparametrized while the first one is rotated [6]. Here the curves are again given on input as discrete sets of nodes in the curves, the same number of nodes in each curve, and both curves now discretized by the same partition of the unit interval (a uniform partition if the first curve is closed). A finite subset of the nodes in the first curve (possibly all of them, possibly one if neither curve is closed) is then selected, which is interpreted to be the set of so-called starting points of the curve, and the redefined L^2 distance is then minimized with an iterative procedure that alternates computations of optimal diffeomorphisms (a constant number of them per iteration for reparametrizing the second curve) with successive computations of optimal rotation matrices (for rotating the first curve) for all starting points of the first curve. As noted in [5], carrying out computations this way is not only more efficient all by itself, but, if both curves are closed, allows applications of the FFT, as demonstrated in [6] for $d = 2$, for computing successively in an even more efficient manner, optimal rotation matrices for all starting points of the first curve. Results from computations with the implementation of the methods above applied on three-dimensional curves of the helix and spherical ellipsoid kind, are presented in [1].

[1] J. Bernal, J. Lawrence, G. Dogan, and C. R. Hagwood. On Computing Elastic Shape Distances between Curves in d-dimensional Space. In review.

[2] J. Bernal, J. Lawrence, and C. R. Hagwood. Fast Dynamic Programming for Elastic Registration of Curves. In *Proceedings of the DIFF-CVML Workshop*, 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada. (2016)

[3] J. Bernal. Shape Analysis, Lebesgue Integration and Absolute Continuity Connections. NISTIR 8217, National Institute of Standards and Technology, 2018.

[4] J. Bernal and J. Lawrence. Characterization and Computation of Matrices of Maximal Trace Over

[16] https://doi.org/10.18434/mds2-2329

Rotations. *Journal of Geometry and Symmetry in Physics* **53** (2019), 21-53.

[5] G. Dogan, J. Bernal, C. R. Hagwood. A Fast Algorithm for Elastic Shape Distances between Closed Planar Curves. In *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Boston, MA. June 2015.

[6] G. Dogan, J. Bernal, C. R. Hagwood. FFT-Based Alignment of 2d Closed Curves with Application to Elastic Shape Analysis. In *Proceedings of the 1st DIFF-CV Workshop*, British Machine Vision Conference, Swansea, Wales, UK. September 2015.

[7] W. Kabsch. A Solution for The Best Rotation to Relate Two Sets of Vectors. *Acta Crystallographica Section A: Crystal Physics* **32**:5 (1976), 922-923.

[8] W. Kabsch. A Discussion of the Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallographica Section A: Crystal Physics* **34**:5 (1978) 827-828.

[9] J. Lawrence, J. Bernal, and C. Witzgall. A Purely Algebraic Justification of the Kabsch-Umeyama Algorithm. *Journal of Research of the National Institute of Standards and Technology* **124** (2019), 124028.

[10] A. Srivastava and E. P. Klassen. *Functional and Shape Data Analysis*. New York, Springer, 2016.

[11] A. Srivastava, E. P. Klassen, S. H. Joshi, and I. H. Jermyn. Shape Analysis of Elastic Curves in Euclidean Spaces *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**:7 (2011), 1415-1428.

[12] S. Umeyama. Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**:4 (1991), 376-380.

[13] G. Wahba. A Least-Squares Estimate of Satellite Attitude. *SIAM Review* **7**:3 (1965), 409.

True Becquerel: A New Paradigm for 21st Century Radioactivity Measurements

Bradley Alpert
Ryan Fitzgerald (NIST PML)
Denis Bergeron (NIST PML)
Svetlana Nour (NIST PML)
Gordon Shaw (NIST PML)
Daniel Schmidt (NIST PML)
Daniel Swetz (NIST PML)
Mike Verkouteren (NIST MML)

Expanding applications of radioactivity in medicine, energy, and security demand quantification of complex mixtures at uncertainty levels that are currently unachievable. This project will enable measurement of absolute activity (Bq) of radionuclide mixtures, avoiding chemical separation, by analysis of the decay heat signature of gravimetric samples embedded within microcalorimeter detectors. This capability consolidates multiple measurements into one, reducing cost and uncertainty. Success will create a primary realization of the Bq for direct assay of real-world samples at NIST and beyond, resulting in faster clinical trials of new radiopharmaceuticals and a faster, expanded nuclear forensics "fingerprinting" method for improved decision making.

NIST Innovations in Measurement Science funding was awarded this year for this project, which combines dispensing, weighing, and placement of mg-quantity solutions, embedding in a transition-edge sensor (TES) microcalorimeter, characterization of high-rate radioactive decay events to generate an energy spectrum, and analysis of the spectrum to quantify constituents of the sample at the level of 0.1 % uncertainty.

The analysis challenges include (1) characterization of detector dynamics, to enable determination of decay energies of events with poor temporal separation, avoiding detector dead time, at an accuracy that reflects the

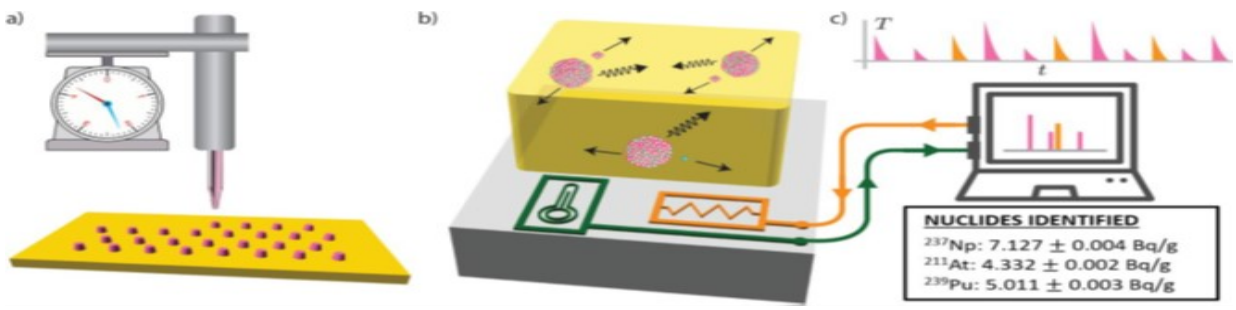


Figure 17. *Schematic: a) Inkjet-dispensing balance deposits mass of radioactive sample into Au absorber. Absorber is folded to create 4π geometry and is bonded to chip. b) Decay products or digital heater (orange) pulses are thermalized creating TES signals. c) Algorithm uses data and model to produce nuclide assay report.*

exquisite precision of the TES detector, (2) characterization of the partial energy losses due to transport out of the absorber material of alpha, beta, and gamma rays, and (3) disambiguation of the spectrum into constituents, based on a library of radionuclide decays, with full quantification. Two new tools for detector dynamics characterization are (a) fabricated capability for electronic excitation of the detector with known energy depositions, and (b) the adjoint sensitivity method, of much recent attention [1], to determine an ODE system from its input/output behavior. This machine learning (ML) technique will be combined with more conventional supervised ML for library-based disambiguation of spectra. The uncertainty, and risk, for both techniques is whether the stringent accuracy requirements of the project can be achieved.

[1] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems* **31** (2018).

TOMCAT: X-ray Imaging of Nanoscale Integrated Circuits for Tomographic Reconstruction

Bradley Alpert
Dan Swetz, Zachary Levine, et al. (NIST PML)
Kurt Larson, et al. (Sandia National Laboratory)
Edward Garboczi (NIST MML)
George Barbastathis et al. (MIT)

As the leading semiconductor manufacturing techniques progress through 14 nm, 10 nm, and now 7 nm technology nodes, the ability to fabricate these chips has outrun the ability to image them. This limitation makes a variety of diagnostic needs much more difficult to satisfy. The NIST Quantum Sensors Group (PML), in collaboration with researchers at Sandia National Laboratory, is leading a project for IARPA's RAVEN (Rapid Analysis of Various Emerging Nanoelectronics) program to develop a small-laboratory capability to image integrated circuits by x-ray tomography. Distinct from other RAVEN projects, TOMCAT exploits a scanning electron microscope (SEM) rather than a synchrotron beamline and does not destroy the chip under test. This is enabled by the exquisite energy resolution of NIST-developed cryogenic microcalorimeter spectrometers, comprised of transition-edge sensors (TES), which are being extended to larger arrays (now to 3000 detectors), as well as with better individual-detector throughput (up to 1000 counts/s) and energy resolution (< 10 eV FWHM). The detectors measure fluorescent photons produced when SEM electrons strike a

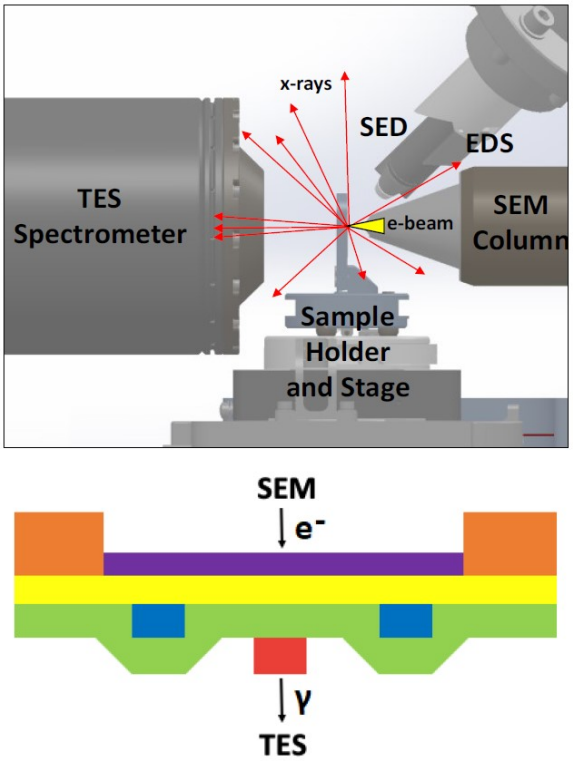


Figure 18. *(Top) layout of SEM chamber and components and (bottom) example target/sample layer stack. During operation, the electron beam of an SEM is focused onto a target layer (purple), which generates x-rays in a localized spot. A spacer layer (yellow) is placed beneath the target layer, which is used to set the magnification and to block excess electrons from penetrating into the sample layers (blue, red, and green). X-rays generated in the target layer can be distinguished from background x-rays due to the TES energy resolution. The intensity of the detected x-rays depends on the attenuation through the sample, which varies as a function of local sample composition and thickness.*

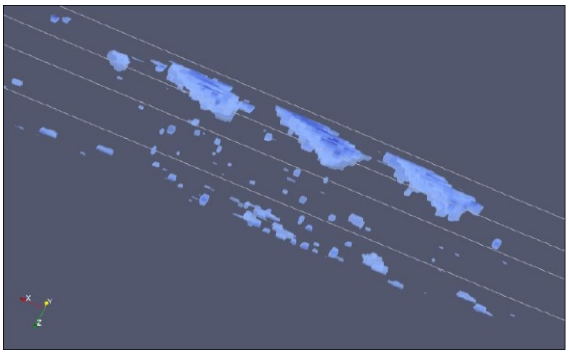


Figure 19. *Three-dimensional image created with maximum a posteriori estimation technique, based on prior NIST work (Z. H. Levine, T. J. Blattner, A. P. Peskin, and A. L. Pintar. Scatter Corrections in X-Ray Computed Tomography: A Physics-Based Analysis. Journal of Research of the National Institute of Standards and Technology 124, 124013, May 2019), to perform tomographic reconstructions including scatter corrections.*

target following their differential attenuation by different materials in the chip

A principal analysis challenge of the project, enabling tomographic structure recovery in this limited-angle, photon-starved regime, is the development of physics-assisted machine learning (PAML) customized for the details of photon fluorescence, absorption, and scattering in this instrument configuration. George Barbastathis, who has had considerable success in PAML for optics, is leading this ML work.

Modeling Magnetic Fusion

Geoffrey McFadden
Eugenia Kim (New York University)
Antoine Cerfon (New York University)

The controlled nuclear fusion of a hot plasma of hydrogen isotopes that is confined by a strong magnetic field is currently an area of intense experimental and theoretical research. Provided that many technological challenges can be overcome, this approach may someday provide a source of commercial energy that avoids many of the drawbacks of conventional nuclear reactors.

We are currently participating in the continuing development of a code, NSTAB, that computes three dimensional equilibria of toroidal plasmas, and determines their nonlinear stability [1, 2]. A crucial assumption is that of "nested" flux surfaces, which describe the geometry of the magnetic field lines that permeate the plasma. This assumption precludes the possibility of magnetic "islands" that would result in undesirably rapid particle transport across magnetic surfaces. A simplified version of NSTAB that applies in a rectangular geometry and avoids the complication of the magnetic axis that occurs in toroidal geometries has recently been developed [3].

To complement the numerical solutions that are produced by this code, we have recently developed an analytic solution to the governing equations for a vacuum magnetic field that does not assume the existence of nested flux surfaces, and which allows the formation of islands. The analytical solution is obtained by starting from a two-dimensional harmonic function that is a periodic Green's function for Laplace's equation, with a smooth set of level surfaces. This solution is analogous to an integrable system such as a two-dimensional pendulum problem in dynamical systems theory. If an appropriate symmetry-breaking, three-dimensional harmonic function is added to the two-dimensional base state, the resulting system is no longer integrable. The level sets then break up into island structures that can be detected numerically by line tracing the magnetic field (given by the gradient of the potential.) An example in

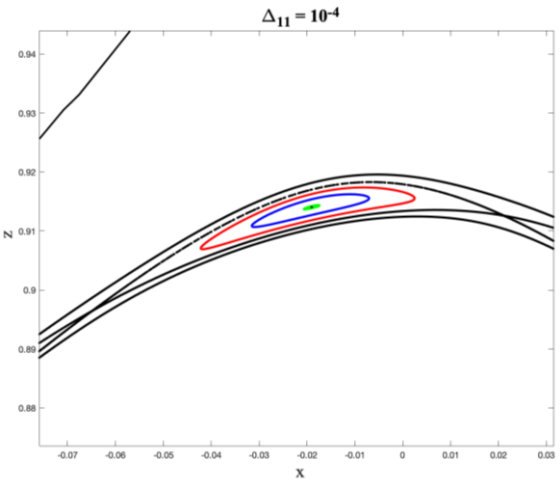


Figure 20. *A Poincare plot of magnetic field lines indicating the formation of magnetic islands. The islands are caused by a symmetry-breaking, three-dimensional perturbation applied to a two-dimensional integrable vacuum field with nested flux surfaces. Here Δ_{11} is the amplitude of the three-dimensional perturbation that produces the islands.*

shown in Figure 20. This analytical solution is being used to generate estimates for the width of islands that can be used as a diagnostic in codes that assume nested flux surfaces and preclude the occurrence of magnetic islands.

[1] M. Taylor. A High Performance Spectral Code for Nonlinear MHD Stability. *Journal of Computational Physics* **110** (1994), 407-418.

[2] P. R. Garabedian and G. B. McFadden. Design of the DEMO Fusion Reactor Following ITER. *Journal of Research of the National Institute of Standards and Technology* **114** (2009), 229-236.

[3] E. Kim, G. B. McFadden, and A. J. Cerfon. Elimination of MHD Current Sheets by Modifications to the Plasma Wall in a Fixed Boundary Model. *Plasma Physics and Controlled Fusion* **62** (2020) 044002.

Optimizing Unlicensed Band Spectrum Sharing with Subspace-Based Pareto Tracing

Zachary J. Grey
Andrew Dienstfrey
Susanna Mosleh (NIST CTL)
Jacob Rezac (NIST CTL)
Yao Ma (NIST CTL)
Jason Coder (NIST CTL)

Motivated by demands for more efficient use of limited radio spectrum, next-generation wireless communications devices are likely to share spectrum between multiple radio access technologies. While spectrum

sharing holds promise, balancing operational requirements of new network paradigms against those of incumbent technologies presents significant technical barriers. To meet these challenges, NIST CTL researchers, in collaboration with ACMD, are developing measurement and simulation techniques to verify configurations of wireless devices which allow for coexistence while minimizing degradation.

In one such example of managing spectrum scarcity, providers are beginning to operate Long-Term Evolution License-Assisted Access (LAA) in unlicensed bands. Operating LAA in unlicensed bands improves spectral-usage efficiency. However, LAA may or may not impede Wi-Fi operation depending on details of operational and environmental variables. Understanding and addressing these challenges calls for a deep dive into the operations and parameter selection of both networks in the medium access control (MAC) and physical layers (PHY).

Mathematically, we formulate spectrum-sharing analysis as a multi-criteria optimization problem in which both Wi-Fi and LAA key performance indicators (KPIs), such as network throughputs on unlicensed bands, are simultaneously maximized with respect to their PHY and MAC layer parameters. This set of maximizing arguments quantifies the inherent trade-off between LAA and Wi-Fi data throughputs. While this is a desirable goal, it is complicated by the high dimensionality of the problem. Furthermore, a systematic parameterization the Pareto front of non-dominated solutions is needed. We briefly discuss our solutions to these problems in the following.

The multi-criteria optimization formalism we propose is complicated by the high dimensionality of the input space. Specifically, our model requires 17 MAC and PHY variables to characterize the coexistence performance. Nevertheless, previous experience suggests that not all variables (or combinations of variables) are equally important in determining qualities of KPIs. We used active subspace analysis to provide a systematic dimension reduction strategy, providing a rigorous foundation for these observations.

In general, given a function $S: \mathcal{D} \subset R^m \to R$, active subspace analysis proceeds by eigendecomposition of the positive definite, $m \times m$, matrix

$$C = \int_{\mathcal{D}} \nabla S(\theta) \nabla S(\theta)^T \, d\mu(\theta).$$

The *active* subspace is defined by the column span of $U_r = [u_1 \ldots u_r] \in R^{m \times r}$ where u_i are the first r eigenvectors associated with the decaying order of eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. For $r < m$, we utilize approximations of the form $S(\theta) \approx H(U_r^T)$ for $H: R^r \to R$. In this way, an r-dimensional model of the original function can be found. A benefit of the active subspace approach is that a rigorous theory exists for bounding the overall approximation error in terms of the eigenvalues of the matrix C defined above. See [1, 2] for details.

Returning to the communications problem, previous work resulted in analytical models for the LAA and Wi-Fi throughputs, indicated by $S_{\mathcal{L}}$ and $S_{\mathcal{W}}$, which are described in detail by Mosleh et al. [3-5]. Both are functions of 17 MAC and PHY parameters which are

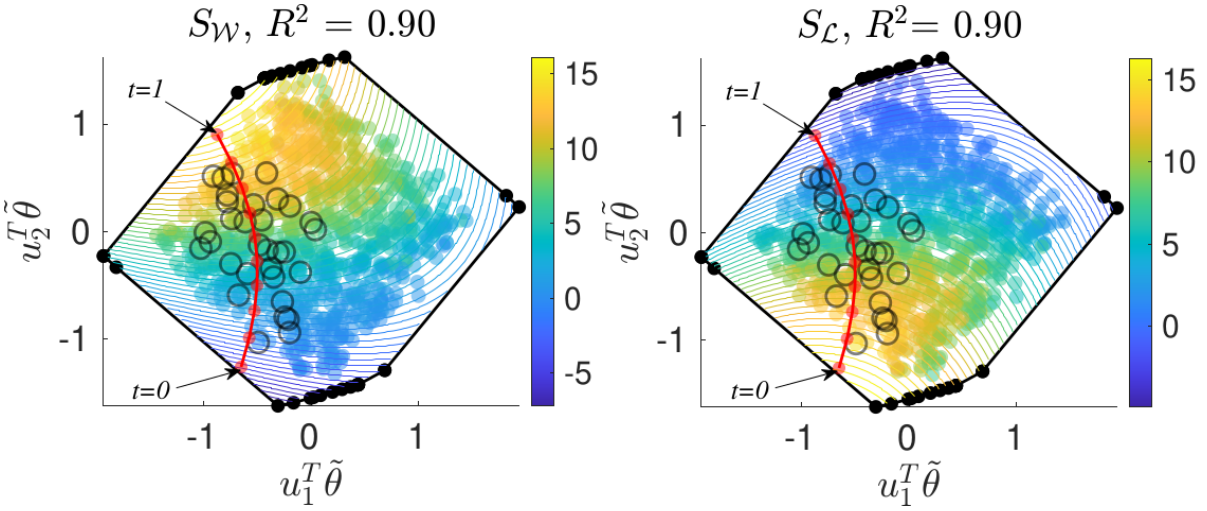


Figure 21. *Pareto trace of quadratic ridge profiles. The quadratic Pareto trace (red curve and dots) is overlaid on a shadow plot over the mixed coordinates (colored scatter) with the projected bounds and vertices of the domain (black dots and lines). Quadratic approximations (colored contours) are computed as least-squares fits over the mixed subspace coordinates. Also depicted is the projection of the non-dominated domain values from 1000 random samples (black circles). The trace begins at* $\boldsymbol{t = 0}$ *with near maximum quadratic Wi-Fi throughput and we move (smoothly) along the red curve to* $\boldsymbol{t = 1}$ *obtaining near maximum quadratic LAA throughput maintaining an approximately best trade-off over the entire curve restricted to a projection of the origin. Note: original parameter scales were transformed for numerical considerations (among other reasons) thus resulting in the accented (tilde) vector.*

collected into a single vector θ. Furthermore, these functions are conditioned on other physical characteristics of the communication network which are collected into a *scenario* vector of fixed values denoted by x. Thus, the complete LAA and Wi-Fi throughput model is denoted by

$$S_{\mathcal{L}}: R^m \times \{x\} \to R: (\theta, x) \mapsto S_{\mathcal{L}}(\theta; x),$$
$$S_{\mathcal{W}}: R^m \times \{x\} \to R: (\theta, x) \mapsto S_{\mathcal{W}}(\theta; x).$$

Active subspace analysis applied to the models $S_{\mathcal{L}}$ and $S_{\mathcal{W}}$ respectively suggests that both may be approximated sufficiently well by functions restricted to two-dimensional subspaces ($r = 2$) in the ambient 17-dimensional space. The issue is that these subspaces are determined by independent eigendecompositions of $C_{\mathcal{L}} = U_{\mathcal{L}}\Lambda_{\mathcal{L}}U_{\mathcal{L}}^T$ and $C_{\mathcal{W}} = U_{\mathcal{W}}\Lambda_{\mathcal{W}}U_{\mathcal{W}}^T$, whose eigenspaces are not the same in general. Fortunately, the space of r-dimensional subspaces in m-dimensional Euclidean space forms a Riemannian manifold known as the Grassmann manifold. Considering each of the subspaces as a point in this space, we utilized the geodesic (i.e., shortest) path connecting these two points, and monitored the errors from constructing reduced-order approximations with respect to a common subspace,

$$S_{\mathcal{L}}(\theta; x) \approx H_{\mathcal{L}}(U(s)^T\theta),$$
$$S_{\mathcal{W}}(\theta; x) \approx H_{\mathcal{W}}(U(s)^T\theta).$$

Here $U(s)$, $s \in [0,1]$, is a point on the geodesic path between $U_{2,\mathcal{L}}$ and $U_{2,\mathcal{W}}$ approximations. As residual errors were monotonic along this path, each improving as the path approached the subspace defined by its independent analysis, the optimal joint subspace was defined as the point along this path where the residual error curves intersected. In this way, a common two-dimensional active subspace was defined for reduced-order approximation of both LAA and WiFi throughputs. In this treatment, the orthogonal complement of the common active subspace defines the *inactive* subspace.

The next problem was to generate a systematic analysis for identifying solutions to the multi-objective optimization problem. We solved this using Pareto tracing techniques described in detail recently by Bolten, et al. [6]. Under this formalism, one defines the *Pareto trace* as the image over the set of solutions to the following optimization problem:

$$\max_{\theta \in \mathcal{D} \subset R^m} tS_{\mathcal{L}}(\theta; x) + (1-t)S_{\mathcal{W}}(\theta; x),$$

for all $t \in [0,1]$ where $\mathcal{D}$ is the domain defined by well-established intervals for each parameter. It turns out that this trace can be solved as an initial-value problem for an ordinary differential equation (ODE). The existence of a 15-dimensional subspace over which the objectives $S_{\mathcal{L}}$ and $S_{\mathcal{W}}$ do not vary entails that this ODE system is unstable. However, the active subspace dimension reduction eliminates this degeneracy. The result is a smooth trajectory, first defined in the active subspace coordinates, and then lifted to a curve $\theta(t)$ through MAC and PHY parameter space. The convex combination of throughputs is maximized over this trace [6].

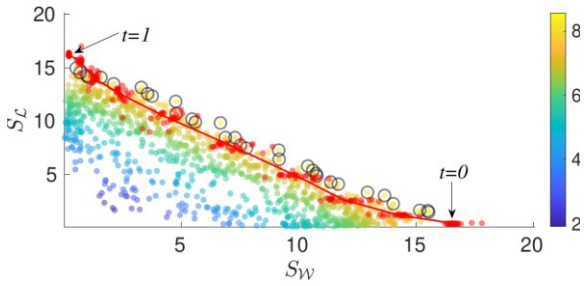


Figure 22. *Approximation of the Pareto front resulting from the quadratic trace. The approximate Pareto front (red curve) is shown with the non-dominated throughout values (black circles) and scatter of 1000 random responses colored according to the averaged throughputs. The red curve is the image of the continuous trace through parameter space (visualized as the red curve in Figure 21 representing a near Pareto optimal set of solutions.*

Observing Figure 21, the continuous Pareto trace over the subspace coordinates (red curve) moves approximately through the collection of projected non-dominated designs (black circles). The non-dominated designs are determined from 1000 random samples of the full 17-dimensional space; sorted according to Kung et al. [7]. However, it is not immediately clear through this visualization that the non-dominated designs constitute elements of an alternative continuous approximation of the Pareto front, perhaps represented by an alternative low-dimensional manifold (*commonly sought in artificial intelligence and machine learning applications*). Instead, we have supplemented a continuous parametrization of the Pareto front which is implicitly regularized as a solution over 2-dimensional subspace coordinates. However, there are infinite θ in the original parameter space which correspond to points along the trace depicted in Figure 21, i.e., infinitely many $m - r$ inactive coordinate values which may change throughputs albeit significantly less than the two mixed active coordinates, $y_1 = u_1^T\theta$ and $y_2 = u_2^T\theta$. To reconcile the choice of infinitely many inactive coordinates, we visualize subsets of 25 inactive coordinate samples drawn conditionally over the inactive subspace (15 orthogonal directions) for each active coordinate occurring over the discretized trace depicted in Figure 21.

Figure 22 depicts the corresponding throughput evaluations along the trace in addition to the conditional inactive samples as red dots along the approximated Pareto front—the red line connects conditional averages of throughputs over inactive samples along corresponding points over the trace, constituting an approximation of the Pareto front. The visualization emphasizes that throughputs change significantly less over the inactive coordinates in contrast to the range of values observed over the active coordinate trace.

Our work has facilitated an approach for simultaneously optimizing network KPIs sharing limited unlicensed spectrum resources. An exploratory analysis utilizing an example of LAA coexistence with Wi-Fi network identified a common subspace-based dimension reduction of a basic model of network behavior. This enabled visualizations and low-dimensional approximations which led to a continuous approximation of the Pareto frontier for the multi-criteria problem of maximizing all convex combinations of network throughputs over MAC and PHY parameters. Such a result simplifies the search for parameters which enable high quality performance of both networks, particularly compared to approaches which do not operate on a reduced parameter space. Analysis of the LAA-Wi-Fi example revealed an explainable and interpretable solution to an otherwise challenging problem, devoid of any known convexity until subsequent exploration. A full paper on this topic has been accepted to the 2021 *Proceedings of the IEEE International Conference on Communications* (ICC).

An extended description of the work will be submitted to *IEEE Transactions on Signal Processing*. The extended manuscript will incorporate alternative low-dimensional approximations including both alternatives of Grassmannian interpolation or subspace unions to improve the trace. We will also summarize a sensitivity analysis, active subspace approximation diagnostics, and an explicit parametrization of a predominantly flat manifold of near Pareto optimal solutions.

Future approaches will enable spectrum sharing in unlicensed bands by simplifying the design of wireless network operation and architecture, ultimately quantifying MAC and PHY parameter combinations with reduced intrinsic dimension giving near-optimal KPI trade-offs. In recent work (currently in review) CTL researchers studied how changing Wi-Fi transmit power impacted KPIs of the networks such as percentage error rate and transmission throughput. They introduced a new concept, the region of wireless coexistence (RWC), representing a set of transmit powers which lead to acceptable values of KPIs. This technique reduces required measurements by around 40 % compared to a baseline uniform measurement. However, scaling to larger problems requires identification the RWC in higher dimension wireless coexistence scenarios. These scenarios will have tens or hundreds of independent variables, leading to a dramatic increase in the volume of the candidate space of RWC sets. Existing RWC surrogate models (Gaussian process) are unlikely to be effective for such high dimension problems. As such, we aim to use the dimensionality-reduction techniques described here to combat issues of computational complexity in larger problems. We hope to accelerate the RWC algorithm by experimenting with how to efficiently take measurements over a manifold of reduced intrinsic dimension, similar to the one identified in this work. We anticipate that the corresponding measurement requirements for estimating RWC will be orders-of-magnitude fewer than with a technique which does not use adaptive sampling in conjunction with dimension reduction.

[1] Z. J. Grey and P. G. Constantine. Active Subspaces of Airfoil Shape Parameterizations. *AIAA Journal* **56**:5 (2018), 2003–2017.

[2] P. G. Constantine. *Active Subspaces: Emerging Ideas in Dimension Reduction for Parameter Studies*. SIAM, Philadelphia, 2015.

[3] S. Mosleh, Y. Ma, J. B. Coder, E. Perrins, and L. Liu. Enhancing LAA Co-existence using MIMO Under Imperfect Sensing. In *IEEE Globecom Workshops*, Waikoloa, HI, 2019, 1-7.

[4] S. Mosleh, Y. Ma, J. D. Rezac, and J. B. Coder. Dynamic Spectrum Access with Reinforcement Learning for Unlicensed Access in 5G and Beyond. In *IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, Antwerp, Belgium ,2020, 1-7.

[5] S. Mosleh, Y. Ma, J. D. Rezac, and J. B. Coder. A Novel Machine Learning Approach to Estimating KPI and PoC for LTE-LAA-Based Spectrum Sharing. *IEEE International Conference on Communications Workshops (ICC Workshops)*, Dublin, Ireland, 2020, pp. 1-6.

[6] M. Bolten, O. T. Doganay, H. Gottschalk, and K. Klamroth. Tracing Locally Pareto Optimal Points by Numerical Integration. Preprint arXiv:2004.10820, 2020.

[7] H.-T. Kung, F. Luccio, and F. P. Preparata. On Finding the Maxima of a Set of Vectors. *Journal of the ACM* **22**:4 (1975), 469–476.

Large Scale Dynamic Building System Simulation

Anthony Kearsley
Amanda Pertzborn (NIST MML)
Aaron Chen (Drexel University)
Jin Wen (Drexel University)

The building sector represents the largest primary energy-consuming sector in the United States, responsible for 41 % of the country's primary energy, in comparison to 28 % for the transportation sector and 32 % for industry. Moreover, buildings consume 74 % of the electricity in the United States, which makes the building sector significant to the overall smart grid infrastructure. Given the rapid development of the smart grid and the potential of buildings to store and generate electricity through demand shifting and transactive control, there is an urgent need to improve the dynamic interactions between buildings and the smart grid, which further calls for robust and accurate dynamic building energy system modeling and simulation.

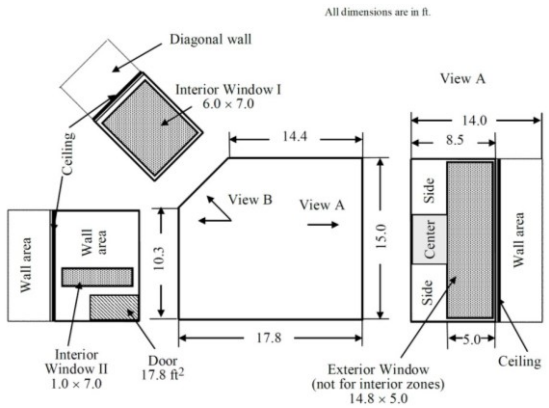


Figure 23. *A floor and window view of a typical zone.*

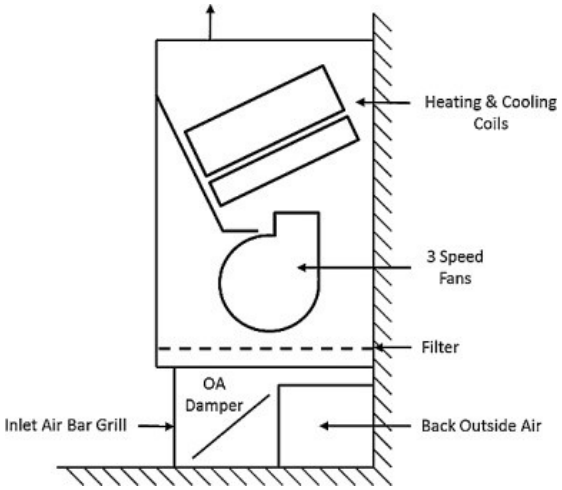


Figure 24. *A simple fan/coil unit schematic.*

Traditional dynamic building simulations typically focus on a single building. However, those used for smart grid applications must simulate large and complex building energy systems and their interactions among building clusters that are composed of multiple buildings. These result in large systems of coupled, potentially ill-conditioned nonlinear systems (easily including thousands of equations), that will need to be solved rapidly. Thus, efficient, robust and accurate solution of large sparse nonlinear algebraic and differential equation systems is becoming more and more essential to meet the demands to simulate large scale multiple building heating ventilation and air conditioning (HVAC) scenarios that are coupled to various complex energy sources either through the smart grid or other means, such as district heating/cooling.

In practice, these simulations are decomposed into geometric zones (see Figure 23) where one expects air flow and thermal conductivity to be similar. Typically, these zones are comprised of rooms or collections of closely connected rooms. HVAC systems control temperature and air flow in these zones through coupled boundary conditions that are matched in what is called a "super-block." It is interesting that something as simple as a fan/coil unit (see Figure 24) can result in an approximation whose discretized system is quite hard to solve. More detailed mathematical models which would more accurately predict the physics of a simple fan coil unit would result in simulations that would be prohibitively expensive. Less detailed mathematical models yield more well-conditioned systems that can be solved quickly but do not provide the flexibility of, for example, more complicated scenarios through various degrees of occupancy, seasons or other factors that must be considered.

In [1], Chen develops a class of large-scale Krylov methods for solving these nonlinear systems using automatically built preconditioners. A numerical survey of the performance of these methods was written this year, [5] and submitted for publication. Comprised of smaller models, like the one used to predict fan/coil dynamics, the resulting large-scale systems are assembled at the same time as a preconditioner, both of which would be made available to a building engineer seeking to simulate and evaluate a building HVAC system. Not surprisingly, poor variable scaling [2] can also lead to a deterioration in numerical performance but results presented in [1] suggest that pre-conditioning of inexact Krylov methods can outperform more traditional dense solution techniques [3] as problem sizes grow. Figure 25 demonstrates that the numerical method proposed in [1], here called INB-PSGMRES(m), outperforms Powell's hybrid method (PH) [4] irrespective of seasonal changes. Here m denotes the dimension of the Krylov subspace, traditionally a parameter that is tuned. Of course, the comparison of CPU time is not exactly a straightforward one. The PH method was not designed for problems of this size. However, it provides strong evidence that methods like those proposed in [1, 5] are very much appropriate for these simulations.

[1] Z. Chen. *Advanced Solver Development for Large-Scale Dynamic Building System Simulation*. Ph.D. Thesis, Drexel University, Philadelphia PA, 2019.

[2] Z. Chen, J. Wen, A. J. Kearsley, and A. J. Pertzborn, Scaling Methods for Dynamic Building System Simulation in an HVACSIM+ Environment. In *Proceedings of the 15th International Building Performance Simulation Association (IBPSA) Conference*, San Francisco, CA, US, 2017, 2059-2065.

[3] S. Pourarian, A. J. Kearsley, J. Wen, and A. Pertzborn. Efficient and Robust Optimization for Building Energy Simulation. *Energy and Buildings* **122** (2016) 53-62.

[4] M. J. Powell. A Hybrid Method for Nonlinear Equations. *Numerical Methods for Nonlinear Algebraic Equations* **7** (1970), 87-114.

[5] Z. Chen, A. Pertzborn, A. J. Kearsley, and J. Wen. Inexact Newton Method for Solving Nonlinear Algebraic Equations in Large-Scale Dynamic Building System Simulation. In review.

Figure 25. *A comparison of required CPU-time for both Powell's Hybrid Method (PH) and the more complicated technique proposed in [1].*

Mathematics of Biotechnology

As proof-of-concept academic work in engineering biology meets the market realities of bringing lab science to product initiation, there are questions in how to compare biological products, measure whether desired outcomes are realized, and optimize biological systems for desired behaviors. NIST is working to deliver tools and standards to measure such biological technologies, outputs, and processes from healthcare to manufacturing and beyond. We support this effort with the development and deployment of innovative mathematical modeling and data analysis techniques and tools.

Metrology for Microfluidics

Paul Patrone
Anthony Kearsley
Amy Q. H. Li (University of Texas at Austin)
Gregory A. Cooksey (NIST PML)
Jalal Sadeghi (NIST PML)
Matthew DiSalvo (NIST PML)

Advances in microfluidics promise to fundamentally change the way many scientific disciplines operate by enabling precise control over the motion of fluids and solutes in micron-scale systems. For example, microfluidic fabrication techniques allow chemical reaction vessels to be reduced to the size of a human hair, parallelized, and subsequently interconnected. This facilitates high-throughput experimentation and combinatorial testing; see Figure 26. Moreover, the medical community has recognized that related approaches can, in principle, enable rapid and simultaneous screening of hundreds of diseases with only a few drops of blood. Indeed, the lure of disrupting this $75 billion medical testing industry has led to significant venture capital investments in high-profile startups that specialize in microfluidics for lab-on-a-chip applications. Despite the promise, few commercial implementations have emerged, leading the community to reflect on the fundamental measurement challenges that have stalled progress.

To address these shortcomings, ITL and PML staff have adopted an approach that tightly integrates mathematical analysis, modeling, physics, and bioengineering to tackle all aspects of the measurement problem simultaneously. Furthering work from FY 2019, several of us continued the development of a pair of flowmeters that can continuously measure volumetric flow rates V_v down to scales that were previously unattainable. These devices expose a fluorescent dye to laser light in the microfluidic channel. After sufficient dosage, the dye photobleaches or degrades; thus, penetration depth of the dye into the laser region is inversely proportional to the fluid velocity. As Figure 27, Figure 28, and Figure 29 illustrate, these devices can be operated in one of two modes by using: (i) symmetry breaking arguments to detect departure from zero flow with a characteristic uncertainty of 0.1 nL/min; and (ii) physics-based scaling relationships to extrapolate below the lower limit of a previously calibrated flow meter, typically to the nL/min range. In past years, this work has led to a pair of patent applications and several publications [1-4].

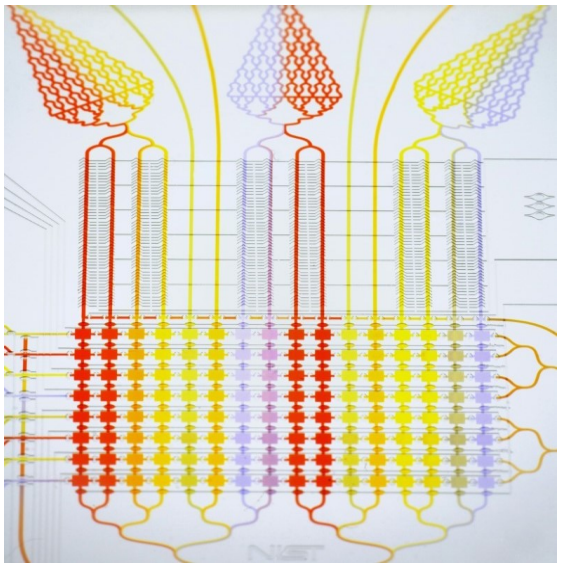


Figure 26. *Example of a microfluidic device used to establish concentration gradients. Fluid of different colors flows through the input channels (top) and enters the experimental chambers (squares). Because the flow is laminar, diffusion is the only mechanism responsible for mixing (which occurs in the upper branching network and middle mixing regions). In this way it is possible to precisely control concentrations of reactants in combinatorial testing assays.*

Recent work has focused on improving the sensitivity and accuracy of our devices by redesigning the underlying optics. In previous designs, rectangular waveguides patterned on the microfluidic chips tended to yield conical laser profiles with diffuse, long tails parallel to the direction of the flow. This diffuse light has the potential to degrade fluorophores before they reach the detection region, effectively reducing the measurable signal. To address this problem, we have redesigned the waveguides to have parabola-like geometries that focus light rays parallel to one another, similar to the operation of a parabolic antenna. This yields well-confined, step-function laser profiles; see Figure 28. Initial experiments with the flowmeters operating in both modes suggest that at least a decade improvement can be obtained with such modifications. Simulation and

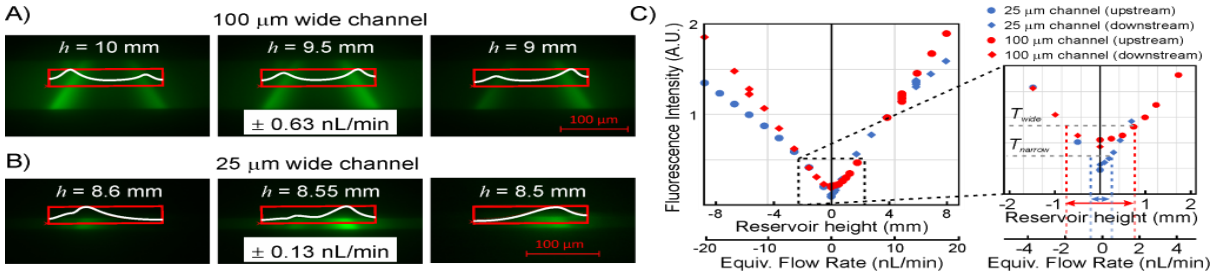


Figure 27. *Determination of uncertainty in height around zero flow. Microscopy images show fluorescence intensity (green) at different heights near zero flow in wide (A) and narrow (B) channel interrogation regions. The white curve shows a line scan of fluorescence intensity across the illumination region (along the bottom line of the red box). When flow is near zero, all fluorescein in the channel is bleached, except for fresh fluorescein diffusing into the edges of the laser path. Brighter intensity on the left is used to indicate positive flow, brighter intensity on the right indicates negative flow. Zero flow is a point somewhere between. (C) Steady-state fluorescence values from waveguides upstream (equivalent to being positioned on the left in the images) and downstream (right in the images) of the excitation beam are shown for the two microchannel widths at different heights of the fluid reservoir(e.g., the flow controller). The equivalent flow rate was determined from the conductance of the system (2.5 nL min-1 mm-1). Inset shows magnification of the critical region and application of thresholds, T_{wide} and T_{narrow} (Gray dotted lines), to isolate the minimum of the signal and determine the height , and corresponding flow, uncertainty (blue dotted lines show 25 um channel; red dotted lines show 100 um channel).*

optimization are on-going to determine the waveguide shapes that yield laser profiles closest to the ideal step function.

We have also developed a new, more advanced theory for data analysis of the flowmeter operation. When operating at non-zero flow, the concentration of fluorophores (which is proportional to the measurement signal) is accurately modeled by a system of integro-differential equations of the form

$$I = \int_{\Omega} dz F(c(z), z)$$
$$c_z = -g(\xi) B(c(z), z)$$

where c is the fluorophore concentration, I is the fluorescence efficiency (fluorescence per input laser power), $\xi = p/v_v$ is the dosage, p is the power, $g(\xi)$ characterizes the dependence of photobleaching on dosage, $B(c(z), z)$ characterizes the rate of photobleaching on concentration and the laser profile shape, and $F(c(z), z)$ is the fluorescence rate. Past results demonstrated that I is a monotone decreasing function of ξ, which provides the scaling relationship used to perform measurements [1]. As a calibration step, however, it is necessary to construct $I(\xi)$ for a known flowrate by varying p and fitting a monotone function to the resulting data. To improve the accuracy of such fitting, we recently proved that under general and physically reasonable conditions, the fluorescence efficiency $I(\xi)$ is also a convex function of dosage; see Figure 29. The main idea behind this proof is to recast the differential equation for c as into a second order partial differential equation (PDE) in terms of dosage. By extending the technique of Piccard iteration to this PDE setting, we prove existence and regularity of the solution and determine the conditions under which it is convex. Results are provided in a recently published manuscript [5].

[1] P. N. Patrone, G. Cooksey, and A. J. Kearsley. Dynamic Measurement of Nanoflows: Analysis and Theory of an Optofluidic Flowmeter. *Physical Review Applied* **11** (2019), 034025.

[2] G. A. Cooksey, P. N. Patrone, J. R. Hands, S. E. Meek, and A. J. Kearsley. Dynamic Measurement of Nanoflows: Realization of an Optofluidic Flow Meter to the Nanoliter-per-Minute Scale. *Analytical Chemistry* **91** (2019), 10713-10722.

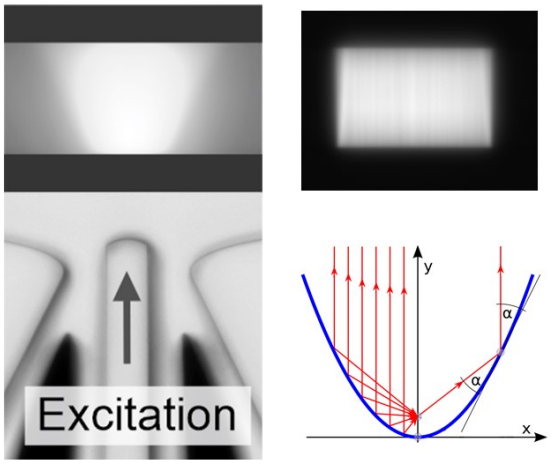


Figure 28. *Comparison of laser profiles.* Left: *an un-optimized waveguide leads to a cone-shaped laser profile. Note also that there is significant diffuse light away from the main cone as indicated by the greyscale.* Right: *A laser profile that uses a focusing strategy akin to a parabolic reflector to achieve a near-rectangular profile. Note the sharp edges in comparison to the conical profile*

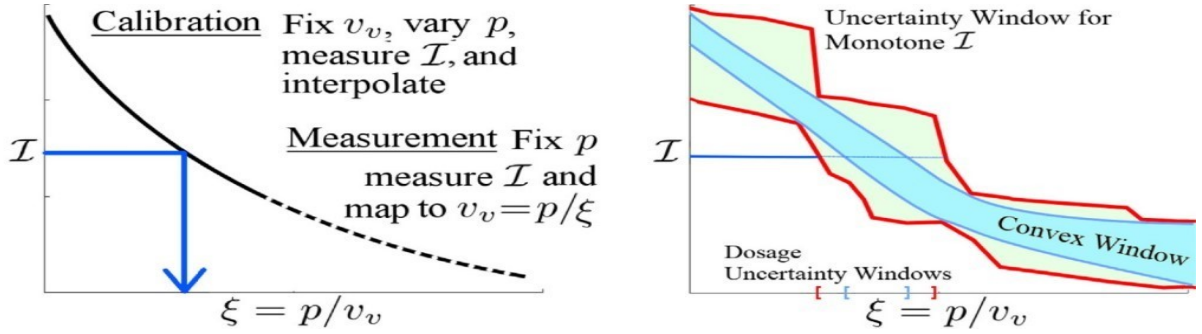


Figure 29. *Process by which the flowmeter is calibrated and used.* Left: *one first fixes a known flow rate and varies the laser power to determine the bijection* $\boldsymbol{I}(\boldsymbol{\xi})$. *Then, the flow rate can be changed, and the laser power fixed. Mapping the fluorescence efficiency to a dosage via the previously determined bijection determines the flow rate.* Right: *Schematic illustrating how uncertainty can be decreased by knowing that* $\boldsymbol{I}(\boldsymbol{\xi})$ *is also a convex function of dosage.*

[3] G. A. Cooksey, P. N. Patrone, J. R. Hands, S. E. Meek, and A. J. Kearsley. Dynamic Measurement of Nanoliter per Minute Flow by Scaled Dosage of Fluorescent Solutions. In *Proceedings of the 22nd International Conference on Miniaturized Systems for Chemistry and Life Sciences* (MicroTas 2018), Kaohisung, Taiwan, November 11-15, 2018.

[4] G. A. Cooksey, P. N. Patrone, and A. J. Kearsley. Optical Flow Meter for Determining a Flow Rate of a Liquid. U.S. Patent Application Number 15/967966, 2017.

[5] P. N. Patrone, A. Q.H. Li, G. A. Cooksey, and A. J. Kearsley. Measuring Microfluidic Flow Rates: Monotonicity, Convexity, and Uncertainty. *Applied Math Letters* **112** (2021), 106694.

A Novel Measurement Method in Microfluidics Via Boundary Layer Theory

Joe Klobusicky
Paul Patrone
Anthony Kearsley

As microfluidic devices become increasingly useful tools for metrology, there is need to develop methods to better characterize their properties, e.g. cross-sectional shapes and the corresponding impacts of flow dynamics. From a practical standpoint, such tasks are challenging due to the small scales involved. Three-dimensional microscopy provides some information but is limited in resolution, while destructive inspection methods cannot be used to characterize devices in use. Thus, there is a need for novel techniques that leverage modeling to extract information about microfluidic devices in non-traditional ways.

ACMD staff have continued work on a project whose main goal is to use boundary layer theory for Poiseuille flow to directly visualize the flow profile, thereby indirectly characterizing the system geometry; see Figure 30. Fluorescent molecules are advected across a channel until reaching a laser, which is sufficiently strong to destroy (or bleach) them on contact. From a modeling standpoint, the quantity of interest is the concentration of unbleached fluorophores $c(r,t)$ at time t at some location $r = (x,y,z)$ inside the channel Ω. The evolution of $c(r,t)$ s given by the advection diffusion equation

$$\partial_t c(r,t) + Qu(x,y)\,\partial_z c(r,t) = D\Delta c(r,t), \quad r \in \Omega.$$

Here D is the diffusivity of unbleached fluorophores, and Q is the volumetric flow rate. The normalized Poiseuille velocity profile satisfies $\int_P Pu(x,y)\,dxdy = 1$, and varies depending on the cross-sectional geometry P of the channel. To simplify the analysis, we consider steady state solutions satisfying

$$Qu(y)\,\partial_z c(r,t) = D\Delta c(r), \qquad r \in \Omega.$$

Instantaneous bleaching of fluorophores produces a Dirichlet condition at the laser interrogation region, and channel walls supply a Neumann condition forbidding fluid to leave. Adjacent to the laser, one can derive a boundary layer solution for the concentration that is accurate up to approximately a distance of $P^{-1} = D/Q$, where P is the nondimensional Peclet number measuring the ratio between drift and diffusion. As of $P^{-1} \rightarrow$

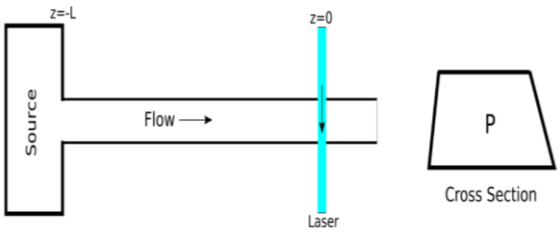


Figure 30. *Schematic of device geometry for flowmeter experiment.* Left: *Fluid beginning at a source with an initial concentration of fluorophores is advected across a channel until reaching a laser interrogation region, where fluorophores are immediately bleached.* Right: *The cross section of the channel, which in general is a trapezoid with unknown angles.*

0, the boundary layer becomes smaller, but the boundary layer solution becomes more accurate, which typical errors of $O(P^{-1})$.

Surprisingly, it is possible to show that the boundary layer solution has level sets whose profiles are the magnitude of the Poiseuille flow velocity as a function of position. The general theory for developing such solutions for elliptic equations is given in [2]. The approximation method works equally well for flow in three-dimensional channels with two-dimensional cross-sectional velocities, even if the cross-section geometry is unknown. This is especially important when considering that device construction often produces channels having trapezoidal cross-sections of unknown dimensions. See Figure 31 or a comparison of estimated velocities and the Poiseuille velocity profile in two dimensions.

A critical part of relating the asymptotic theory to real-world application is the determination of appropriate parameter regimes in which the boundary layer estimates are feasible, and for which device construction is also within the capabilities of laboratories. With new capabilities developed at NIST for measuring slow flows [1], the boundary layer method for measuring velocity is possible under a reasonable collection of fluorophore concentrations and channel length dimensions. For future research, we hope to verify this theory and simulations with lab experiments, and also to develop estimates from uncertainty quantification which address potential points of error such as image resolution and diffusion of laser light near the interrogation region.

[1] G. A. Cooksey, P. N. Patrone, J. R. Hands, S. E. Meek, and A. J. Kearsley. Dynamic Measurement of Nanoflows: Realization of an Optofluidic Flow Meter to the Nanoliter-Per-Minute Scale. *Analytical Chemistry* **91** (2019), 10713-10722.

[2] J. Kevorkian. *Partial Differential Equations: Analytical Solution Techniques.* Texts in Applied Mathematics **89871**, Springer, 1990, 456-7.

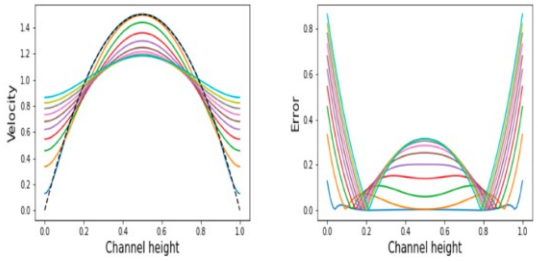


Figure 31. Left: *Comparison of parabolic one-dimensional Poisseuille velocity profile (dotted line) and variation of velocity estimates as* $\boldsymbol{P^{-1}}$ *ranging from .005 to .3. As* $\boldsymbol{P^{-1}}$*decreases, velocity estimates better approximate the Poisseuille velocity profile.* Right: *Percent error between estimated velocity and Poisseuille velocity profile.*

Thermodynamics of DNA Origami

Paul Patrone
Anthony Kearsley
Robert DeJaco
Alex Liddle (NIST PML)
Jacob Majikes (NIST PML)

Understanding the fundamental physical properties of DNA is a critical problem in bioengineering. Beyond being a basic building block of life, DNA is increasingly finding applications as a material that can be used to build bio-compatible microstructures. Indeed, the study of *DNA origami*, the nanoscale folding of DNA to create arbitrary two- and three-dimensional shapes at the nanoscale, has the potential to revolutionize biomanufacturing by leading to new paradigms that leverage powerful and robust phenomena such as self-assembly [1]. Despite this, measurement techniques to characterize the thermodynamic properties of DNA are hampered by the challenges associated with the study such small systems: variation in sample preparation, the need for indirect measurements, and large background effects that decrease signal-to-noise ratios.

To address these problems, ACMD and PML staff are collaborating on a project whose main goal is to develop analysis and uncertainty quantification (UQ) tools for widely used measurement techniques in the study of DNA origami. Initial work has focused on measurement of melt-curves, which characterize the degree to which double-stranded DNA separates into two individual strands as a function of temperature. Such measurements are useful for extracting thermodynamic information and can also be used to detect mutations.

A typical measurement uses a polymerase chain-reaction (PCR) machine to generate data. The protocol uses fluorophores that predominantly emit light when the DNA is in a double-stranded state. Thus, ramping the temperature changes the relative fraction of single and double-stranded DNA, which subsequently affects the fluorescence signal. In typical systems, however, small sample sizes and pipetting errors make it virtually impossible to repeat experimental conditions, so that appreciable differences in measurement signals are common. Background effects, especially when they are temperature dependent, add further complications.

To overcome this lack of reproducibility, we developed a data analysis and UQ strategy that can determine when and to what extent there is a "universal signal" connecting a collection of similarly prepared samples. The underlying intuition behind this approach is the recognition that differences in sample sizes (e.g., due to pipetting errors) should only affect the magnitude of signal, but not the physics that generated it. Thus, we postulate that all signals $S_i(T)$ can be related by an equation of the form

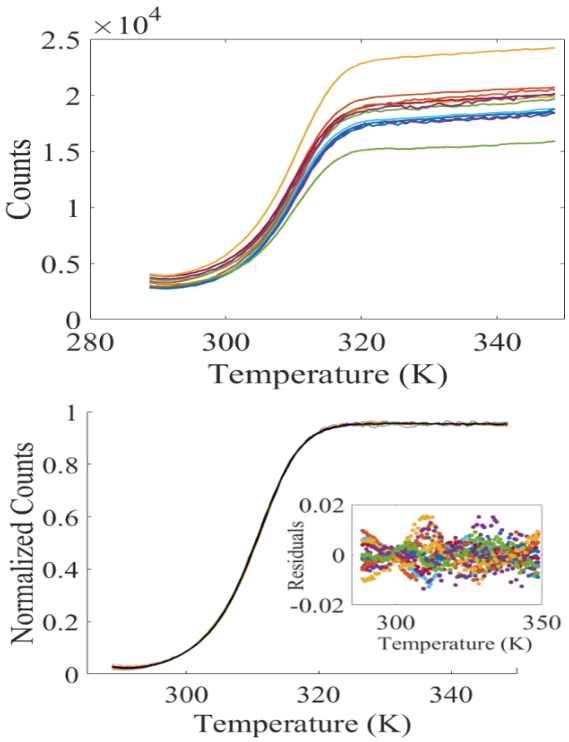


Figure 32. *Raw PCR data (top) and transformed data (bottom). Despite the fact that the raw data exhibits O(1) relative variation, the transformed signals all agree to within a few percent pointwise in temperature after data collapse.*

$$U(T) = \tau_{0,i} S_i(T) + \sum_{n=1}^{N} \tau_{n,i} B_n(T) \qquad (1)$$

where i is measurement index, $U(T)$ is the universal (or scaled) form of the signal, the $B_n(T)$ are background corrections, and the $\tau_{n,i}$ are affine parameters that depend on the measurement signals and bring them all into agreement with the universal signal. To determine the $\tau_{n,i}$, we minimize an objective function of the form

$$L = \sum_{i,j,T} (U_i(T) - U_j(T))^2$$

where $U_i(T)$ is the realization of $U(T)$ generated from $S_i(T)$ via an affine transformation in the spirit of Eq. 1.

Note that this objective considers all combinations of transformed signals, although it requires regularization and the addition of constraints to ensure a well-posed problem. Additional constraints, e.g. formulated in terms of relative errors, can be used to test the feasibility of achieving collapse with a given uncertainty in the universal signal. The bottom subplot of Figure 32 illustrates the result as applied to the collection of datasets in the top sub-plot; see also [1] and [2].

Recent work has focused on extending these techniques to characterize the kinetics of competitive DNA interactions. In one such example, a DNA loop can be induced by an oligomer "staple" to close, effectively forming a double-stranded state. However, binding of multiple staples to the DNA can inhibit ring-closing, which thus has an appreciable effect on the fluorescence signal. By varying the relative fractions of staples to DNA, it is then theoretically possible to quantify their binding kinetics through the low-temperature signals. However, low signal-to-noise ratios and baseline offsets make it difficult to identify the zero signal, which is necessary for quantifying the fraction of rings that cannot close because of competitive staple binding. To address this problem, we modified the procedure described above to provide Monte Carlo estimates of zero, based on a statistical analysis of samples without DNA. Results of this analysis are described in a recently accepted manuscript [4]. Current work is focused on extending our results to and widely used class of intercalating dyes, which have additional background effects in the presence of single-stranded DNA.

[1] J. M. Majikes and J. A. Liddle. DNA Origami Design: A How-To Tutorial. *Journal of Research of the National Institute of Standards and Technology* **126** (2020), 001.

[2] P. N. Patrone, A. J. Kearsley, J. M. Majikes, and J. A. Liddle. Analysis and Uncertainty Quantification of DNA Fluorescence Melt Data: Applications of Affine Transformations. *Analytical Biochemistry* **607** (2020), 113773.

[3] J. M. Majikes, P. N. Patrone, D. Schiffels, M. Zwolak, A. J. Kearsley, S. Forry, and J. A. Liddle. Revealing Thermodynamics of DNA Origami Folding via Affine Transformations. *Nucleic Acids Research* **48** (2020), 5268–5280.

[4] J. Majikes, P. N. Patrone, A. J. Kearsley, M. Zwolak, and J. A. Liddle. Failure Mechanisms in DNA Self-assembly: Barriers to Single Fold Yield. *ACS Nano* **15** (2021) 3284–3294.

Metrology for Cytometry

Paul Patrone
Anthony Kearsley
Geoffrey McFadden
Danielle Middlebrooks
Matthew Roberts
Gregory Cooksey (NIST PML)
Matthew DiSalvo (NIST PML)
Jalal Sadeghi (NIST PML)
Sumona Sarkar (NIST MML)
Lili Wang (NIST MML)

For more than 30 years, flow cytometry, a technique used to measure characteristics of cells, has been a mainstay for cancer detection, drug development, and biomedical research. It has remained a primarily qualitative metrology platform, however, because measurement uncertainties associated with this technique are so large. While exact economic figures are

difficult to estimate, this has clearly had a significant impact on the roughly $200 billion of waste in the healthcare industry and contributed to the broader reproducibility crisis in biomedical research [1]. The challenge of making cytometry an accurate and precise metrological tool arises from the competing requirement that it have high throughput. Typical biological samples can have up to hundreds of millions of cells, which must be analyzed over a few hours.

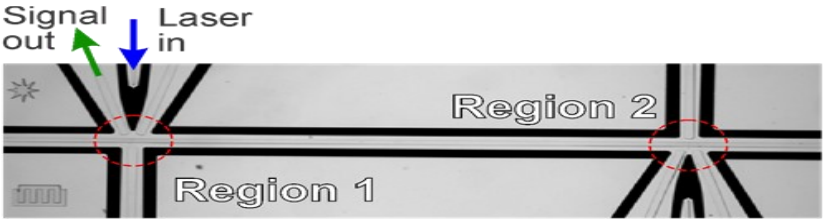


Figure 33. *Microscope image of a cytometer under development. Particles flow from left to right in the central channel. At each interrogation regions, an excitation laser causes objects to fluoresce. The signal – light collected as a function of time – is further analyzed to deduce properties of the objects. The use of separate measurement regions allows for repeat measurements to characterize uncertainty.*

To achieve this throughput, cytometers direct cells through a microfluidic channel at high-speed, past an optical interrogation region that collects fluorescence light from antibodies attached to surface proteins. The total fluorescence collected from each cell should then, in principle, be proportional to the total number of markers on its surface. But in practice, this idealized picture is complicated by the cumulative effects of the physical phenomena involved: fluid-dynamic forces cause cells to move across streamlines and/or have unpredictable trajectories; optical geometric collection efficiencies depending on position in the interrogation region; and signal acquisition and processing tools introduce nonlinear effects and measurement uncertainties through discrete sampling. These challenges, in addition to the complexity of exactly replicating the necessary measurement infrastructure at the micron scale, have made it virtually impossible to reproduce measurements on a single cell, a necessary first step towards fully assessing and controlling uncertainties in cytometry.

ACMD, PML, and MML staff were awarded a NIST Innovations in Measurement Science (IMS) award to develop a microfluidic-based cytometer whose design explicitly allows control and study of repeat measurements of cells; see Figure 33. Following work in FY 2019, we began testing a new hydrodynamic focusing strategy that deposits particles on off-center inertial nodes, which are equilibrium streamlines for objects in flow. Concurrently, we have been developing a classification strategy that leverages optimal decision theory to match objects in different regions; see Figure 34. Interestingly, this method uses reproducibility (e.g., in the velocity of an object) to define the probability of a measurement outcome at a downstream interrogation region, conditioned on an event upstream. Amazingly, we find that particle identification between two regions can be performed with a typical accuracy of roughly 99.9 %, with more advanced classification methods yielding up to an additional decade. These results are the subject of a pair of manuscripts to be submitted in FY 2021. Moreover, we anticipate having the world's most accurate cell counter by this time next year.

In addition to this, we have initiated a series of research projects on uncertainty quantification (UQ) for cytometry. The first of these aims to identify the relative contributions to the total uncertainty budget arising from preparation of samples, variations due to instrument operators, and randomness in properties of reagents (which are complex biomarkers). The main idea behind our approach is to: (i) formulate a mathematical model that describes the physical interplay between these effects during the measurement process; and (ii) apply likelihood analyses to experimental data to characterize the relative contributions of each effect. This work is summarized in a publication prepared jointly with FDA and Fluidigm [2] and has motivated additional UQ work associated with SARS-CoV-2 antibody testing [3].

The second UQ project aims to characterize the history-dependent behavior of background effects in a cytometer. Empirical evidence suggests that debris from previous measurements can pollute a measurement at

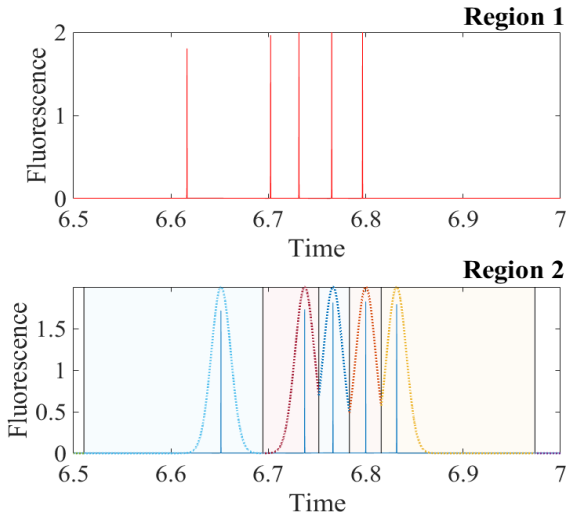


Figure 34. *Schematic of the classification strategy used to match events in region 2 with those in region 1; see also Figure 33. Given the probability density of the time-of-flight, we construct a partition (shaded intervals) of times in region 2. Each interval in the partition corresponds to a single event in region 1 and thus forms the basis for classification. For example, the event occurring in the light-blue domain of the region 2 time-series is identified with the first event in region 1; the event in the light-red domain corresponds to the second event in region 1, etc.*

later times, e.g., if debris is trapped in the system for some time. This cross-contamination leads to false signals that confound data analysis. To address this, we are developing a modeling framework wherein a measurement signal can be modeled via an equation of the form

$$S(r,t) = n_q(t)Q(r) + \sum_i n_i(t)B_i(r)$$

where $S(r,t)$ is the probability of a measurement outcome r as a function of time t, $Q(r)$, and $B_i(r)$ are the probabilities of measurements conditioned on the sources being the target measurand (Q) or one of many possible background sources (B_i), and the $n(t)$ are time-dependent normalization factors characterizing the rate of events associated with any given population. The main idea is to characterize the time-dependence of the background effects by sequentially measuring increasingly complicated samples, e.g., water, followed by beads, single-cell samples, double-cell samples, etc. Having characterized the time-dependence of the first m effects, it should be possible to subtract their relative contributions from a signal generated when characterizing the $(m+1)^{\text{th}}$ effect. Work to validate the usefulness and accuracy of this framework is on-going.

In addition to this, the IMS project is pursuing several related technologies. Work is in progress to develop an amplitude-modulation cytometer that can use a single detector to collect all measurement signals, which can be demultiplexed via signal processing. This has led to a provisional patent application [4]. We have also been pursuing related projects on signals filtering and device optimization. Future work aims to introduce live cells and perform repeat measurements of their properties.

[1] W. H. Shrank, T. L. Rogstad, and N. Parekh. Waste in the US Health Care System: Estimated Costs and Potential for Savings. *JAMA* **322**:15 (2019), 1501-1509.

[2] L. Wang, R. Bhardwaj, H. Mostowski, P. N. Patrone, A. J. Kearsley, J. Watson, L. Lim, J. Pichaandi, O. Ornatsky, D. Majonis, S. Bauer, and H. Degheidy. Establishing B-cell Reference Control Materials for Comparable and Quantitative Cytometric Expression Analysis. *PLoS ONE* **16**:3 (2021), 1-18.

[3] L. Tian, E. Elsheikh, P. N. Patrone, A. J. Kearsley, A. Gaigalas, S. Inwood, S. Lin-Gibson, D. Esposito, and L. Wang. Towards Quantitative and Standardized Serological and Neutralization Assays for COVID-19. *International Journal of Molecular Sciences* **22**:5 (2021), 2723.

[4] G. A. Cooksey, A. J. Kearsley, and P. N. Patrone. Multiplexed Amplitude Modulation Photometer and Performing Multiplexed Amplitude Modulation Photometry. NIST Docket No. 20-037CIP1, U.S. Non-Provisional Patent Application.

Mathematical Models for Cryobiology

Daniel Anderson
Anthony Kearsley
James Benson (University of Saskatchewan, Canada)

Cryobiology, the study of biological specimens at cryogenic temperatures, plays an enormous role in a wide range of fields. In the field of medicine, cryobiology is the basis for cryopreservation in assisted reproduction, organ transplantation, biobanking and personalized medicine. Cryo-banking is used in the agriculture industry as well as for initiatives aimed at preserving rare and endangered plant and animal species and in the development of more productive agricultural yields. Applications in forensics arise in the processing and preservation of frozen biological samples that are often important and fragile evidence in criminal investigations. The breadth and depth of these applications reflect the complexity of the biological, chemical, and physical aspects required to describe and model these problems. Mathematical and computational models can be used to probe these complex systems and, in conjunction with sophisticated control and optimization schemes, can establish more effective protocols for cryopreservation.

Cryopreservation of a human cell is a form of biomimicry that attempts to do in the laboratory, or in silica, what some frog and amphibian species can do naturally. A cryoprotectant, effectively an antifreeze, aims to play the role that substances like glucose play in the winter-preservation of frozen frogs. These cryoprotectants, which are added to the extra-cellular environment, help to remove water from a cell before cooling, thus reducing the likelihood of intracellular ice formation. This, however, comes at a price due to the cell's limited ability to withstand the elevated levels of chemicals, including the naturally occurring salts in the cell, possibly damaging the cell or causing death by chemical toxicity. Thus,

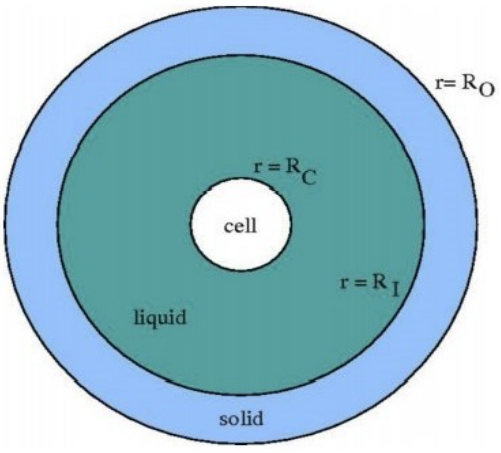


Figure 35. *Spherical cell inside a liquid-solid structure.*

the maintenance of a viable cell during cryopreservation is complicated by two primary factors. Cooling the cell too quickly increases the likelihood of intracellular ice formation and cell damage or death while cooling the cell too slowly can overexpose the cell to high solution concentrations and lead to chemical toxicity.

Mathematical modeling in cryobiology thus requires a detailed understanding of thermal and chemical transport in bulk phases as well as across a semipermeable cell membrane. Additionally, phase transformation of these multicomponent solutions – phenomena that link cryobiology to a wide range of other fields from geophysics to industrial materials processing – must also be included. Cryopreserving a cell requires a delicate balance between two competing damage mechanisms, and thus a delicate optimization problem. In this project we are exploring foundational aspects of biochemical and physical modeling in cryobiology, computational methods for the solution of these models, and applications of these ideas to cryopreservation of cells.

The focus of [1] was to establish the foundations of the chemical thermodynamics necessary to describe transport processes during cryopreservation. This work formulated chemical potentials and related thermodynamics quantities for non-dilute and non-ideal multicomponent solutions of experimental and theoretical interest to cryobiologists. Next, the multiphase, multi-species transport equations were developed along with a consistent characterization of cell membrane dynamics and solid-liquid phase transitions [2]. A critical aspect of our work was to obtain, from first principles, mathematical models that address both spatial and temporal dynamics of chemical species and heat transport. Various aspects of the freezing of a spherical biological cell (see Figure 35) were addressed in [3] and [4]. Based on a numerical algorithm outlined in [4], we explored in [3] the evolution of the thermal fields in the solid, liquid and intracellular regions along with the concentration of cryoprotectant and the intra- and extracellular salts (see Figure 36). These studies incorporated the effects of confinement and partial solute rejection, which had not previously been examined in cryobiology. These observations led to the development of objective functions in [4] which provide a measure of both intracellular undercooling and chemical toxicity. These control functions appear more suited to deriving cooling protocols than previously employed toxicity functions [5].

The mathematical models and computational algorithms established in this foundational series of papers [1-4] will be the launching point for a control and optimization study to develop improved cryopreservation cooling strategies. The central role played by phase transformation of multicomponent systems along with the underlying mathematical descriptions links these cryobiological processes to related ones that occur under vastly different conditions in geophysics and industrial materials processing [6]. Further cryobiological modeling efforts will build on this knowledge to address warming and/or melting phenomena along with the associated optimal protocols.

[1] D. M. Anderson, J. D. Benson, and A. J. Kearsley. Foundations of Modeling in Cryobiology I: Concentration, Gibbs Energy, and Chemical Potential Relationships. *Cryobiology* **69** (2014), 349-360. DOI: 10.1016/j.cryobiol.2014.09.004

[2] D. M. Anderson, J. D. Benson, and A. J. Kearsley. Foundations of Modeling in Cryobiology II: Heat and Mass Transport in Bulk and at Cell Membrane and Ice-liquid Interfaces. *Cryobiology* **91** (2019), 3-17. DOI: 10.1016/j.cryobiol.2019.09.014

[3] D. M. Anderson, J. D. Benson, and A. J. Kearsley. Foundations of Modeling in Cryobiology III: Heat and Mass Transport in a Ternary System. *Cryobiology* **92** (2020), 34-46. DOI: 10.1016/j.cryobiol.2019.09.013

[4] D. M. Anderson, J. D. Benson, and A. J. Kearsley. Numerical Solution of Inward Solidification of a Dilute Ternary Solution Towards a Semi-permeable Spherical Cell. *Mathematical Biosciences* **316** (2019), 108240. DOI: 10.1016/j.mbs.2019.108240

[5] J. D. Benson, A. J. Kearsley, and A. Z. Higgins. Mathematical Optimization of Procedures for Cryoprotectant Equilibration using a Toxicity Cost Function. *Cryobiology* **64** (2012), 144-151. DOI: 10.1016/j.cryobiol.2012.01.001

[6] D. M. Anderson, P. Guba, and A. J. Wells. Mushy Layer Convection. In review.

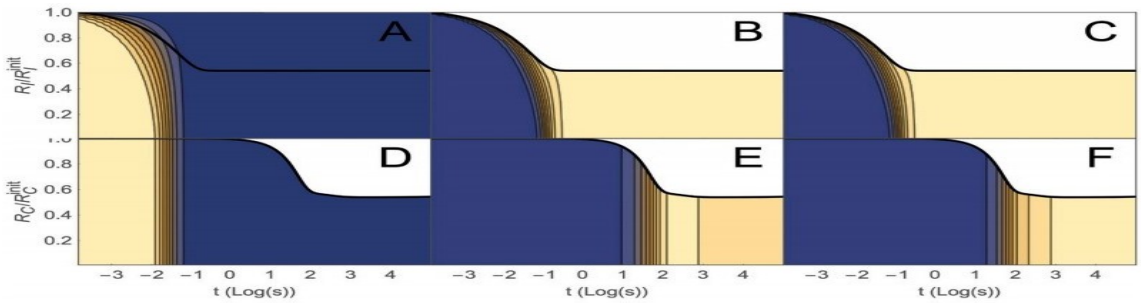


Figure 36. *Solid-liquid interface evolution (bold solid curve, plots A,B, and C) and cell radius evolution (bold solid curve, plots D, E, and F) along with the solid and liquid phase temperature, liquid CPA concentration and salt concentration (plots A, B, and C, respectively) and the intracellular temperature, CPA concentration and salt concentration (plots D, E, and F, respectively).*

Model Performance Predicting Pure Triglyceride Thermodynamic Properties

Anthony J. Kearsley
Arun S. Moorthy (NIST MML)
Julia Seilert (Technische Universität Berlin, Germany)
Eckhard Flöter (Technische U. Berlin, Germany)

Consisting of a glycerol backbone with three fatty acids attached (see Figure 37), triglycerides provide structure and function to vegetable and animal fats which play an important role in food science and engineering. Melting and solidification of fats in multi-component systems directly impacts product characteristics such as melting range and solid-phase composition. These compounds are also a primary element of stores of adipose. Triacylglycerols form an important part of the human diet serving as an energy source when stored in fat tissues. They provide thermal and mechanical protective layers surrounding important organs. Their primary source in the human diet are plant oils.

There is a significant interest in predicting how these fats melt, as this has a tremendous effect on storage/transportation and commercial viability of food products that contain fats (i.e., almost all food products). Thus, building mathematical models that accurate predict fat melting on the product development cycle can have a significant economic impact on the food.

Together with Technische Universität Berlin, we have begun investigating mathematical model fits, performance, and predictive power when imposing thermodynamic constraints on the predictions. Two approaches to ensure thermodynamic consistency were evaluated. Firstly, a parameter set obtained by using a constrained optimization procedure enforcing physically reasonable predictions was evaluated. Secondly, model parameters were derived by optimization to a sub-dataset containing exclusively thermodynamically sound data, that is, data that already satisfied thermodynamic constraints. The dataset we employed contained data spanning nearly three decades of food science research, originating from various references. This led to a dataset of variable quality. An interesting subset we also examined, the well-studied subset of mono-acid saturated TAGs, offered another thermodynamic constraint that predictions should possibly satisfy. This constraint arises from the evolution of incremental enthalpy of fusion and entropy of fusion per carbon. The effect this constraint has on the model performance is currently being evaluated.

Regrettably, our study was cut short, in part due to the pandemic. We plan to continue this research in 2021.

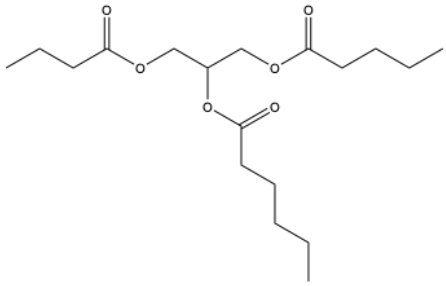


Figure 37. *An example structure drawing of a triglyceride molecule; the most common constituents in dietary fats (e.g. butter, margarines, oils).*

Mass Spectral Similarity Mapping Applied to Fentanyl Analogs

Anthony J. Kearsley
W. Gary Mallard (NIST MML)
Arun S. Moorthy (NIST MML)
William E. Wallace (NIST MML)

Compound identification is a fundamental task in forensics chemistry. A common tool of this process is mass spectral library searching. That is, given a database of mass spectra of known compounds and a mass spectrum of an unidentified compound (analyte), mass spectral library searching seeks to identify the analyte by calculating similarities/dissimilarities [6]. Typically, mass spectral library searching returns a “hit list” or list of likely compound matches. Ideally, top hits will provide an analyst adequate information to correctly infer the identity of an analyte.

It is important to note that the eventual classification of the analyte is still a human task—the burden of identification resides with the analyst. This is of particular interest in forensic chemistry applications. For example, incidents of opioid abuse continue to grow, leading to a rise of fentanyl and related analogs (see Figure 38) with fast onset and high therapeutic indexes [1]. Forensic practitioners struggle to provide confident identifications when encountering novel designer fentanyl analogs.

To aid in this activity, a natural extension to traditional mass spectral library searching is being developed that, in addition to returning a hit list of database entries with similar spectra to the analyte spectrum, generates a map of spectral similarity between the hit list mass spectra themselves. The map can then be scrutinized using a variety of numerical techniques. The objective of this extension is to provide analysts with additional information which can improve confidence in identifying

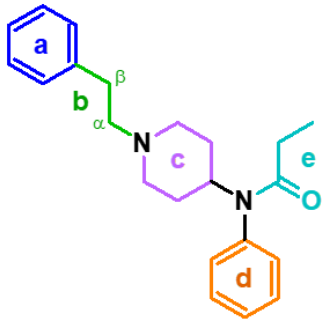


Figure 38. *Molecular structure of fentanyl with potential sites for modification (as interpreted by DEA guidelines [2]) labeled.*

analytes, and, may eventually lead to automated classification with quantifiable uncertainty.

Fentanyl analogs can be classified by the type and location of the structural modifications by which they differ from a fentanyl molecule. The defined modification sites and structural scaffold are an interpretation derived from the definitions provided in [2]. One interesting notion is that of fentanyl analog type indicating the number of structural locations (modification sites) that an analog differs from the molecule fentanyl. For example, α-methyl fentanyl is considered a Type I fentanyl analog, as it differs from fentanyl at a single modification site. Type II analogs have modifications in two locations, and so forth for Types III-V. The spectra and structure information for all Type I fentanyl analogs, along with the spectrum for the molecule fentanyl, contained in the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Mass Spectral Library version 3.3 [3] can be used to form a reference set for comparisons.

Preliminary results suggest the notion of using a map rather than just a hit list can be very useful in analysis. Using Type I fentanyl as an example, a reference set can be generated. In an attempt to classify compounds, the hybrid similarity match factors [3] are exclusively used to approximate spectral similarity when generating maps together with multidimensional scaling (MDS) [4], which is a procedure for representing dissimilarity among pairs of objects as distances between points in a low-dimensional space while preserving correlations from the original data as best as possible. Using MDS to project the Type I fentanyl analog dissimilarity matrices down to two dimensions, one can visualize more easily as seen in Figure 39, which illustrates the spectral similarity space of the Type I fentanyl analog reference set using non-metric MDS.

[1] J. B. Morrow, J. D. Ropero-Miller, M. L. Catlin, A. D. Winokur, A. B. Cadwallader, J. L. Staymates, S. R. Williams, J. G. McGrath, B. K. Logan, M. M. McCormick, K .B. Nolte, T. P. Gilson, M. J. Menendez, and B. A. Goldberger. The Opioid Epidemic: Moving Toward an Integrated, Holistic Analytical Response. *Journal of Analytical Toxicology* **43**:1 (2019), 1-9.

[2] *Federal Register* **83**:25, February 6, 2018, 5188–5192.

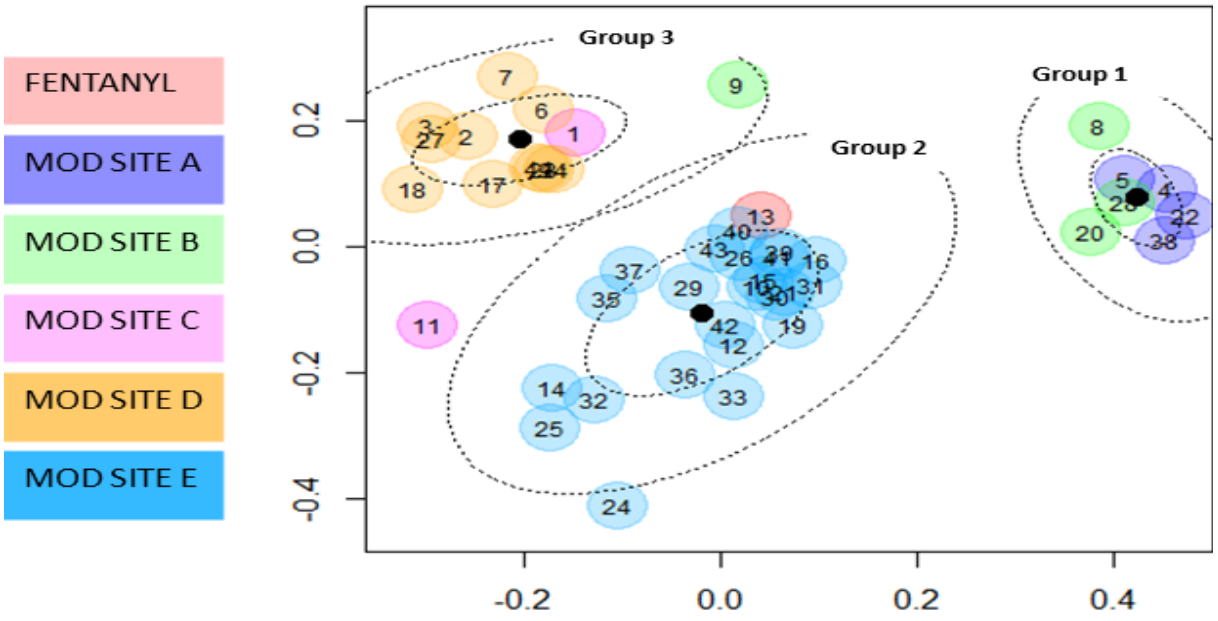


Figure 39. *Spectral similarity space of the Type I fentanyl analog reference set visualized by non-metric multidimensional scaling of dissimilarity matrices generated by hybrid match factors. Each point in the spectral similarity space represents a spectrum of a molecule and its coloring indicates at which modification site it differs from fentanyl (labeled 13, in red). Groups 1-3 were discovered through k-means clustering of the spectral similarity space data, with bold black dots indicating cluster centers and dotted outlines indicating the 50 % (inner) and 95 % (outer) confidence ellipse around each center.*

[3] SWGDRUG Mass Spectral Library version 3.3. URL https://swgdrug.org

[4] A. S. Moorthy, W. E. Wallace, A. J. Kearsley, D. V. Tchekhovskoi, and S. E. Stein. Combining Fragment-Ion and Neutral-Loss Matching during Mass Spectral Library Searching: A New General-Purpose Algorithm Applicable to Illicit Drug Identification. *Analytical Chemistry* **89**:24 (November 20, 2017), 13261-13268.

[5] A. J. Kearsley, R. A. Tapia, and M. W. Trosset. The Solution of The Metric STRESS and SSTRESS Problems in Multidimensional Scaling Using Newton's Method. *Computational Statistics* **13**:3 (1998), 369–396.

[6] A. S. Moorthy and A. J. Kearsley. Pattern Similarity Measures Applied to Mass Spectra. In *Progress in Industrial Mathematics: Success Stories: The Industry and the Academia Points of View* (M. Cruz, C. Parés, and P. Quintela, eds.) ICIAM 2019, Valencia Spain. 43-54.

Predicting Molecular Information from Mass Spectra

Anthony J. Kearsley
Arun S. Moorthy (NIST MML)
Gary Mallard (NIST MML)
William E. Wallace (NIST MML)
Stephen E. Stein (NIST MML)

Measuring the characteristics of individual molecules or compounds is fundamental to modern industrial and applied chemistry. One of the most commonly employed measurement techniques is mass spectrometry, where molecules or compounds are converted into ions that are controlled with external electric and magnetic fields [1]. Typically, these ions are then detected electronically, resulting in the output of a mass spectrum, most commonly visualized as a bar graph, in which each bar represents an ion having a specific mass-to-charge ratio (m/z) and the height of the bar or peak indicating relative ion abundance. One piece of information often sought from mass spectra is an estimate of molecular mass.

Recently we developed and compared three different computational methods for predicting the nominal molecular weight of a compound from a mass spectrum. The first of the three methods compared was the so called "Simple Method" which follows a set of rules to interpret spectral peaks. It requires only a spectrum in order to produce an approximate nominal molecular weight. The second method referred to as the "Hitlist Method" uses the results of a mass spectral library search [2] to generate a molecular weight prediction with an assigned probability. This method requires a spectrum and a reference spectral library to estimate a nominal molecular weight. A final method, known as the "Iterative Hybrid Search Method," executes several hybrid library-searches [3, 4] of a given mass spectrum, before it generates a molecular weight prediction and an assigned probability.

In our study we observed that the hitlist method (86 % success rate) outperformed the simple (67 %) and iterative hybrid (68 %) for correct identifications in a global assessment. Of course, each method requires different amounts of information to be employed and there are important applications, like forensic chemistry, where one method may be more appropriate than another. Some results are summarized in Figure 40 where the left portion reports results from the hitlist method and the right on the iterative hybrid search method for a collection of 10410 test spectra from [5].

[1] J. T. Watson and O. D. Sparkman. *Introduction to Mass Spectrometry: Instrumentation*, Applications and Strategies for Data Interpretation. John Wiley & Sons, 2007.

[2] S. E. Stein and D. R. Scott. Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *Journal of the American Society for Mass Spectrometry* **5**:9 (1994), 859-866. DOI: 10.1016/1044-0305(94)87009-8

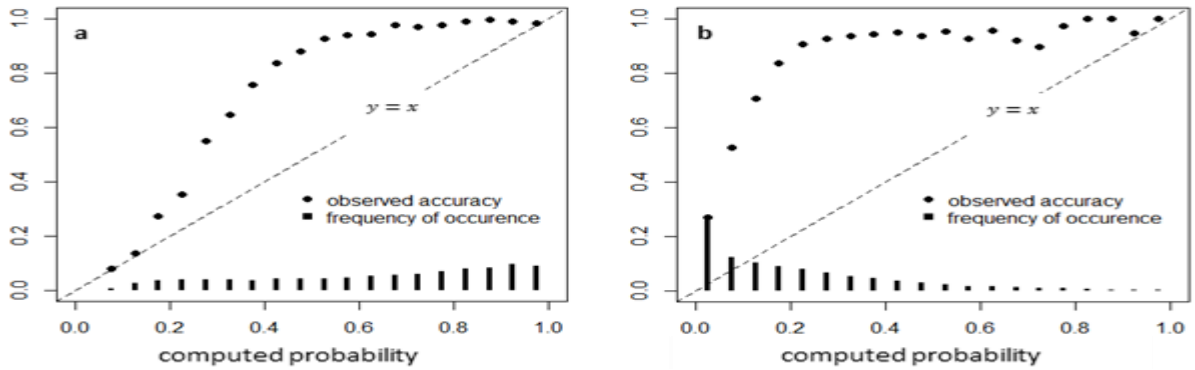


Figure 40. *Summary of performance results for molecular weight prediction algorithms: Hitlist Method (left) and Iterative Hybrid Search Method (right). The Hitlist Method predicted molecular weights with computed probability greater than 0.8 (3628 examples) were correct more than 99 % of the time. The Iterative Hybrid Search predicted molecular weight with computed probability greater than 0.8 (113 examples) were accurate ~99 %.*

[3] A. S. Moorthy, W. E. Wallace, A. J. Kearsley, D. V. Tchekhovskoi, and S. E. Stein. Combining Fragment Ion and Neutral-Loss Matching during Mass Spectral Library Searching: A New General-Purpose Algorithm Applicable to Illicit Drug Identification. *Analytical Chemistry* **89**:4 (2017), 13261-13268. DOI: 10.1021/acs.analchem.7b03320

[4] M. C. Burke, Y. A. Mirokhin, D. V. Tchekhovskoi, S. P. Markey, J. L. Heidbrink Thompson, C. Larkin, and S. E. Stein. The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics. *Journal of Proteome Research* **16**:5 (2017), 1924-1935. DOI: 10.1021/acs.jproteome.6b00988

[5] S. E. Stein. NIST/EPA/NIH Mass Spectral Library (NIST 17) and NIST Mass Spectral Search Program (Version 2.3) User Manual. 2017.

[6] A. S. Moorthy, A. J. Kearsley, W. G. Mallard, W. E. Wallace and S. E. Stein. Inferring the Nominal Molecular Mass of an Analyte from Its Electron Ionization Mass Spectrum. In preparation

Modeling for Biological Field-Effect Transistor Measurements

Ryan M. Evans
Anthony Kearsley
Arvind Balijepalli (NIST PML)

The ability to tailor therapies to individuals or specific subsets of a population to deliver personalized care has the potential to fundamentally improve healthcare delivery. The most promising therapeutic candidates for such targeted care are new classes of biologic drugs based on naturally occurring molecules, made possible due to rapid advances in genomics and proteomics [1]. Importantly, such therapies can be safer and yield better outcomes at lower doses when treating debilitating conditions such as diabetes, Alzheimer's disease, or certain cancers [2]. Unfortunately, widespread use of personalized care is currently limited by our ability to routinely measure pathology in individuals, including biomarkers, metabolites, tissue histology, and gene expression. Moreover, existing clinical diagnostics are cumbersome, require specialized facilities, can take days to weeks to perform, and are in many cases prohibitively expensive.

To overcome such problems researchers have developed new portable detection tools, including antibody-based lateral flow assays [3], microelectromechanical sensor (MEMS) based resonators that can detect binding of biomarkers to the sensor surface [4], surface plasmon resonance [5], ring cavity resonators [6], and electronic measurements with biological field effect transistors (Bio-FETs) [7–9]. The latter are particularly well-suited for biomarker measurements due their high charge sensitivity and direct signal transduction, allowing label-free measurements at physiological concentrations. Also, by leveraging semiconductor processing techniques, measurements with FETs can be made massively parallel, cost-effective, and portable.

A Bio-FET is a three-terminal device as represented in Figure 41. A semiconductor channel between the source and drain terminals conducts a current that is strongly modulated by an electrostatic potential applied to the gate. Biomarkers in aqueous solution exhibit a well-defined electrostatic surface potential [10] arising from charged hydrophilic residues that interact with water. When these molecules adsorb to receptor sites confined to the Bio-FET's floor, they strongly modulate the channel current proportionally to the magnitude of their surface potential. This phenomenon allows Bio-FETs to be used to detect and quantify adsorbed biomarkers in solution. Furthermore, functionalizing the Bio-FET, by attaching molecules to the gate surface that have a high inherent affinity for biomarkers of interest, allows measurements with high specificity.

An accurate time-dependent model that couples transport dynamics to kinetic processes at the surface is needed to identify key parameters associated with Bio-FET experiments, such as kinetic coefficients and diffusion constants. Although Bio-FETs measure the time-dependent change in current, and are thus dynamic experiments in nature, most previous modeling efforts assume a steady distribution of biomolecules immobilized to the Bio-FET's floor and focus on characterizing

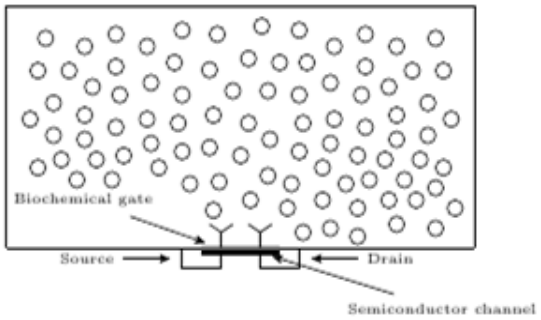


Figure 41. *Schematic of a Bio-FET experiment.*

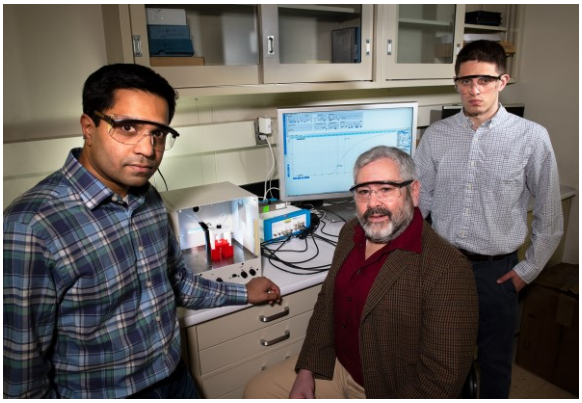


Figure 42. *Schematic of a Bio-FET experiment.*

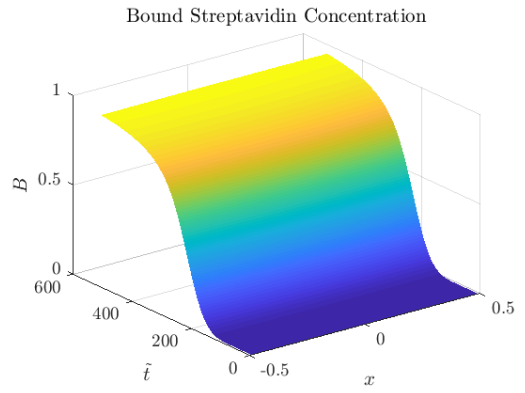


Figure 43. *Space-time curve of reacting species concentration. Here x represents the dimensionless spatial variable and* $\tilde{\mathbf{t}}$ *represents the dimensional temporal variable.*

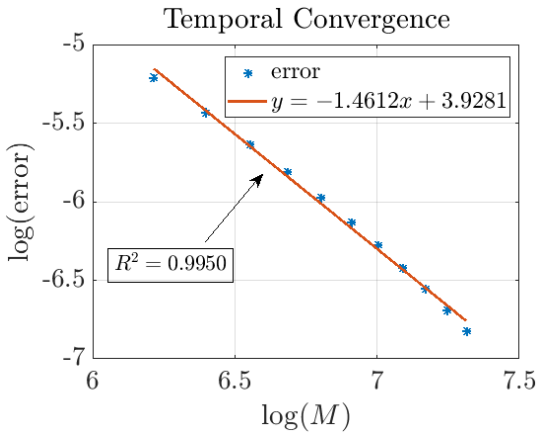


Figure 44. *Error in infinity norm as a function of the number of points in time M.*

charge transport through the semiconductor [11, 12]. Such models are useful for elucidating semiconductor physics but cannot utilize time-dependent measurements to estimate parameters associated with Bio-FET experiments. One could ostensibly use steady-state data to estimate binding affinities as Edwards discusses in [13] in the context of surface plasmon resonance biosensor experiments, but this requires multiple experiments to obtain accurate estimates and yields only the ratio of the association and dissociation rate constants. A dynamic model can utilize measured time-series data from a single experiment to estimate not only kinetic coefficients, but also other important parameters such as diffusion coefficients.

We presented a time dependent model for Bio-FET experiments in [14], which quantified the evolution of the reacting species concentration in presence of a continuous and uniform injection of ligand at the top boundary. Recently, we have developed a mathematical model for a sealed experiment in which a drop of ligand is injected at an instant of time [15]. This model takes the form of a diffusion equation coupled to a nonlinear equation that describes the evolution of the reacting species concentration. Through a Laplace transform and tools from complex analysis, this coupled set of equations has been reduced to a single nonlinear integro-differential equation (IDE) in terms of the reacting species concentration. Though this equation exhibits a singular convolution kernel that approaches infinity at a rate proportional to $1/\sqrt{t}$ as t approaches 0, a numerical solution to this equation has been developed which achieves greater than first-order accuracy. See Figure 43 for a space-time curve that depicts the numerical IDE solution, and Figure 44 for strong evidence of convergence at a rate of $O(\Delta t^{1.46})$ in time, despite the singular convolution kernel.

To compare the numerical solution of the IDE with experimental data, stochastic regression was employed to separate signal from noise in Bio-FET measurements [16]. This involved modeling the Bio-FET signal as a stochastic differential equation that has a deterministic term and a stochastic term:

$$dX_t = (a_0(t) + a_1(t)X_t)dt + b_0(t)dW_t\,,$$
$$X_t(0) = 0.$$

The coefficients $a_0(t)$, $a_1(t)$, and $b_0(t)$ were determined using a local weighted regression and maximum likelihood estimation as described in [16]. Figure 45 shows an overlay of the measured signal and the estimated deterministic component of the signal $X_t = (a_0(t) + a_1(t)X_t)\,dt$. Parameters were estimated using a trust region method implemented as lsqnonlin() in MATLAB. The results are shown in Figure 46, which demonstrates that our IDE model exhibits excellent agreement with experimental data.

Current work centers on optimal design. The signals shown in Figure 45 and Figure 46 are functions of design variables, such as the biochemical gate radius, solution-well radius, height of the solution-well, and receptor concentration. To maximize diffusive flux into the surface it is desirable to make the receptor concentration and biochemical gate radius as large as possible, but this comes at the cost of increasing a signal distortion effect discovered in [15]. The presence of other design variables such as the solution-well radius and height further complicate this highly nonlinear problem. To reconcile these two competing objectives, we are investigating formulating the problem using the Van Stakelberg approach described by Lions [17]. Following the ideas of [18], our future work will study competitive binding kinetics at the surface through a PDE-constrained optimization approach.

[1] K. Fosgerau and T. Hoffmann. Peptide Therapeutics: Current Status and Future Directions. *Drug Discovery Today* **20** (2015), 122-128.

[2] B. K. Binukumar, Y.-L. Zheng, V. Shukla, N. D. Amin, P. Grant, and H. C. Pant. TFP5, a Peptide Derived from

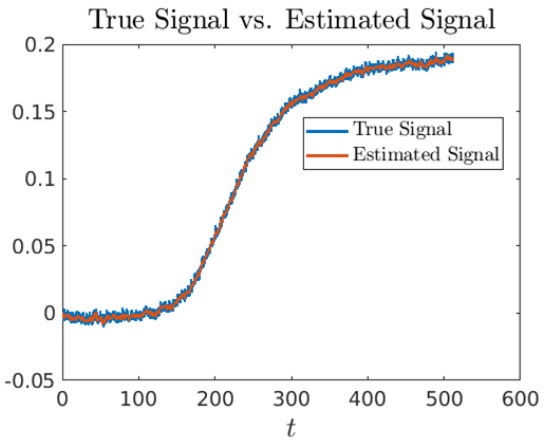


Figure 45. *An overlay of the measured signal (labeled true), and the estimated deterministic component of the signal found through stochastic regression (estimated signal).*

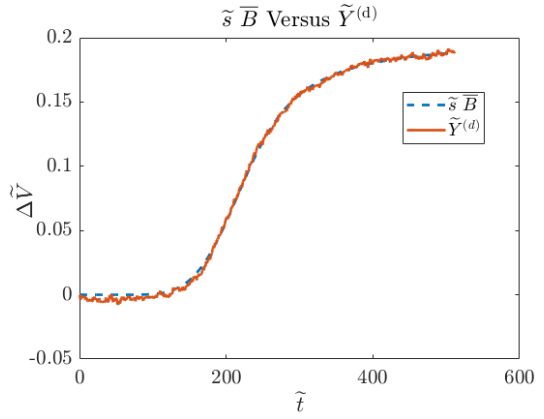


Figure 46. *An overlay of the spatially averaged solution to our IDE model together with the deterministic component of the signal.*

P35, a CDK5 Neuronal Activator, Rescues Cortical Neurons from Glucose Toxicity. *Journal of Alzheimer's Disease* **39**:4 (2014), 899909.

[3] P. K. Drain, L. Gounder, F. Sahid, and M.-Y. Moosa. Rapid Urine LAM Testing Improves Diagnosis of Expectorated Smear-Negative Pulmonary Tuberculosis in an HIV-Endemic Region. *Scientific Reports* **6** (2016), 19992.

[4] B. Ilic, H. G. Craighead, S. Krylov, W. Senaratne, C. Ober, and P. Neuzil. Attogram Detection Using Nanoelectromechanical Oscillators. *Journal of Applied Physics* **95:7** (2004), 3694-3703.

[5] W. Knoll. Interfaces and Thin Films as Seen by Bound Electromagnetic Waves. *Annual Review of Physical Chemistry* **49**:1 (1998), 569-638.

[6] D. K. Armani, T. J. Kippenberg, S. M. Spillane, and K. J. Vahala. Ultra-high-Q Toroid Microcavity on a Chip. *Nature* **421** (2003), 925-928.

[7] Y. Cui, Q. Wei, H. Park, and C. M. Lieber. Nanowire Nanosensors for Highly Sensitive and Selective Detection of Biological and Chemical Species. *Science* **293** (2001), 1289-1292.

[8] P. Mohanty, Y. Chen, X. Wang, M. K. Hong, C. L. Rosenberg, D. T. Weaver, and S. Erramilli. Field Effect Transistor Nanosensors for Breast Cancer Diagnostics. arXiv:1401.1168 [q-bio.QM], 2014.

[9] F. Pouthas, C. Gentil, D. Cote and U. Bockelmann. DNA Detection on Transistor Arrays Following Mutation-specific Enzymatic Amplification. *Applied Physics Letters* **84**:9 (2004), 1594-1596.

[10] A. Cardone, H. Pant, and S. A. Hassan. Specific and Non-Specific Protein Association in Solution: Computation of Solvent Effects and Prediction of First-Encounter Modes for Efficient Configurational Bias Monte Carlo Simulations. *The Journal of Physical Chemistry B* **117**:41 (2013), 12360-12374.

[11] S. Baumgartner, M. Vasicek, A. Bulyha, N. Tassotti, and C. Heitzinger. Analysis of Field-Effect Biosensors Using Self-Consistent 3D Drift-Diffusion and Monte-Carlo Simulations. *Procedia Engineering* **25** (2011), 40700410.

[12] C. Heitzinger, N. J. Mauser, and C. Ringhofer. Multiscale Modeling of Planar and Nanowire Field-effect Biosensors. *SIAM Journal on Applied Mathematics* **70**:4 (2010), 1635–1654

[13] D. A. Edwards. Estimating Rate Constants in a Convection-Diffusion System with a Boundary Reaction. *IMA Journal on Applied Mathematics* **63**:1 (1999), 89-112.

[14] R. M. Evans, A. Balijepalli, and A. J. Kearsley. Diffusion-Limited Reaction in Nanoscale Electronics. *Methods and Applications of Analysis* **26**:2 (2019), 149-166.

[15] R. M. Evans, A. Balijepalli, and A. J. Kearsley. Transport Phenomena in Field Effect Transistors. *SIAM Journal on Applied Mathematics* **80** (2020), 2586–2607.

[16] A. J. Kearsley, Y. Gadhyan, and W. E. Wallace. Stochastic Regression Modeling of Chemical Spectra. *Chemometrics and Intelligent Laboratory Systems* **139** (2014), 26-32.

[17] J. L. Lions. Hierarchic Control. *Proceedings Mathematical Sciences* **104**:1 (1994), 295-304.

[18] A. J. Kearsley, J. W. Tolle and P. T. Boggs. Hierarchical Control of a Linear Diffusion Equation. In *Large-Scale PDE-Constrained Optimization*. Lecture Notes in Computational Science and Engineering **30** (2003). Springer, Berlin, Heidelberg.

Asymptotic Analysis of an Integral Kernel in a Mathematical Model for Biological Field Effect Transistors

Ryan M. Evans
Anthony Kearsley

Precision medicine is an innovative approach to healthcare that involves tailoring therapies to individuals or specific subsets of a population and is widely recognized as a superior alternative to the one-size-fits-

all model which can fail to account for significant differences between patients. The former offers superior treatment at lower doses when treating conditions such as Alzheimer's disease and certain cancers [1]. Although precision medicine shows great promise, a persistent hurdle to widespread adoption has been a lack of low-cost and accurate biomarker measurements. A class of novel medical diagnostic instruments known as biological field effect transistors (Bio-FETs) are being developed by engineers at NIST, and promise to overcome these difficulties by offering low-cost, accurate, and portable measurements on a rapid time scale. This technology could extend accessibility of personalized care domestically, as well as to the challenging medical treatment landscape in developing countries where low-cost and portable point-of-care diagnostics are desperately needed to accurately diagnose debilitating diseases like tuberculosis.

These instruments possess a cylindrical geometry that exhibits azimuthal symmetry, and a two-dimensional cross-sectional schematic of the instrument has been depicted in Figure 41. During a typical experiment, chemical reactants are injected at the top of a solution-well. The reactants then diffuse throughout the well, onto the surface to bind with other chemical reactants immobilized to a region known as the biochemical gate, which physically connects the receptors to a semiconductor channel. Kinetic interactions on the biochemical gate modulate current flow through the semiconductor, resulting in a time-dependent signal such as the one shown in Figure 46.

An accurate time-dependent mathematical model is necessary to make critical diagnostic decisions on a quantitative basis, and until recently such models were such models were unavailable. This problem was addressed when Ryan Evans and Anthony Kearsley of the ACMD teamed up with Arvind Balijepalli to develop the first class of dynamic mathematical models consistent with device physics. The researchers first developed a mathematical model for an experiment in which chemical reactants are injected continuously and uniformly along the top boundary, and it was shown that this model affords insight into counter-intuitive instrument dynamics that are extremely difficult to observe experimentally [4]. Evans, Balijepalli, and Kearsley then decided to model an experiment in which chemical reactants are injected during an instant in time, rather than in a continuous and uniform fashion. The latter scenario is far more complex, because asymptotic analysis used in the former scenario can no longer be used to simplify the problem. Mathematically, this means finding the solution to a dynamic nonlinear partial differential equation (PDE) that contains a discontinuous boundary condition. Evans, Kearsley, and Balijepalli were able to solve this problem by reformulating the governing equations in terms of an integrodifferential system:

$$\frac{\partial B}{\partial t} = (1-B)C(x,0,t) - KB, \tag{1a}$$

$$B(x,0) = 0, \tag{1b}$$

where

$$C(x,0,t) = C_i(x,0,t) - \frac{\mathrm{Da}\ D\ l_s^2}{\epsilon}\Big\{ \int_{-1/2}^{1/2}\int_0^t \theta_3(e^{-(\pi l_s/\epsilon)^2 D\tau})\frac{\partial B}{\partial \tau}(\nu, t-\tau)\ \mathrm{d}\tau\ \mathrm{d}\nu + \sum_{n=1}^{\infty} 2\int_{-1/2}^{1/2}\int_0^t e^{-\lambda_n D\tau}\theta_3(e^{-(\pi l_s/\epsilon)^2 D\tau})\frac{\partial B}{\partial \tau}(\nu, t-\tau)\ \mathrm{d}\tau \cdot \cos(\sqrt{\lambda_n}(\nu+1/2l_s))\ \mathrm{d}\nu \cdot \cos(\sqrt{\lambda_n}(x+1/2l_s))\Big\}, \tag{1c}$$

and $C_i(x,0,t)$ satisfies the Neumann problem on a rectangle subject to appropriate initial data. All of the constants appearing in (1) are dimensionless and positive, and their descriptions are as follows: D is a dimensionless diffusion coefficient, representing the ratio of the diffusive time scale to the reaction time scale; ϵ is the aspect ratio, representing the solution-well height to diameter; l_s is the ratio of the biochemical gate diameter to the solution well diameter; and K the equilibrium dissociation rate constant, which represents the ratio of the time scale for dissociation to the time scale for association. The parameter Da is the Damköhler number, and this key parameter that represents the reaction velocity to diffusion velocity. The magnitude of the *Damköhler* number determines whether kinetic dynamics on the surface are reaction-limited or diffusion-limited. The $\lambda_n = (n\pi l_s)^2$ are eigenvalues associated with an eigenvalue problem in Laplace transform space. We invite the reader to consult [3] for a more detailed discussion and numerical values.

A significant challenge to finding a numerical approximation to (1) is posed by the singular convolution kernel

$$\theta_3\left(e^{-(\pi l_s/\epsilon)^2 D\tau}\right) = 1 + 2\sum_{k=1}^{\infty} e^{-(\pi l_s/\epsilon)^2 \tau}, \tag{2}$$

which may be identified as a third-order Jacobi theta function [2, Eq. 20.2.3]. An examination of the behavior at the origin shows that this series diverges as t approaches zero from the right. Although this does not pose a problem analytically, a form more amenable to numerical discretization is found by applying Poisson's summation formula [2, Eq. 1.8.14] to show

$$\frac{Dl_s}{\epsilon}\theta_3\left(e^{-(\pi l_s/\epsilon)^2 D\tau}\right) = \frac{\sqrt{D}e^{-\lambda_n D\tau}}{\sqrt{\pi t}}\theta_3\left(e^{-\epsilon^2/(l_s^2 Dt)}\right) \tag{3}$$

Since the right-hand-side of (3) diverges at a rate proportional to $1/\sqrt{t}\ as\ t$ approaches zero from the right, it may be shown that by substituting (3) into (1c) that this equation may be regularized and a method of lines approximation may be developed that enjoys outstanding convergence rates of $O(\Delta t^{3/2})$ in time and $O(\Delta t^2)$ in space. The $O(\Delta t^{1/2})$reduction from the usual $O(\Delta t^2)$ directly results from the inverse square root behavior of the convolution kernel.

While (3) shows that the left-hand-side behaves like $1/\sqrt{t}$ as t approaches zero from the right, this may not obvious *a priori*. We can elucidate this matter by using the Mellin transform [2, Eq. 1.14.32] to develop an asymptotic expansion of the left-hand-side of (3). Indeed, taking a Mellin transform of the singular component of the left-hand-side of (3) yields

$$\mathcal{M}\left[\sum_{k=1}^{\infty} e^{-(k\pi l_s \epsilon^{-1})^2 Dt}\right] = \frac{\Gamma(s)\zeta(2s)}{[(\pi l_s \epsilon^{-1})^2 D t]^s} \qquad (4)$$

for $\mathrm{Re}(s) > 1/2$, where the gamma function and Riemann zeta function are defined as in [2, Eq. 5.2.1] and [2, Eq. 25.2.1] respectively. An asymptotic expansion is found by integrating along a large rectangular contour in the complex plane that encloses N poles of (4), applying Cauchy's residue theorem, and substituting the resulting expression into the left-hand-side of (3) to obtain

$$\theta_3\left(e^{-(\pi l_s/\epsilon)^2 Dt}\right) = \sqrt{\frac{D}{\pi t}} e^{-\lambda_n D\tau} + o(t^N) \qquad (5)$$

as t approaches zero from the right. It is interesting to note that this expansion is asymptotic only, which poses no problem analytically [5]. Since the right-hand-side of (3) possess an asymptotic expansion identical to (5) as t approaches zero from the right, these two expressions exhibit the same asymptotic behavior at the origin.

These findings are currently being prepared for publication. There are still many interesting questions associated with Bio-FETs to address, and among them include optimal design and using stochastic regression to find optimal window sizes.

[1] B. K. Binukumar, Y.-L. Zheng, V. Shukla, N. D. Amin, P. Grant, and H. C. Pant. Tfp5, a Peptide Derived from p35, a cdk5 Neuronal Activator, Rescues Cortical Neurons from Glucose Toxicity. *Journal of Alzheimer's Disease* **39**:4 (2014), 899–909.

[2] F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. *NIST Digital Library of Mathematical Functions*. Release 1.0.28 of 2020-09-15. URL: https://dlmf.nist.gov/

[3] R. M. Evans, A. Balijepalli, and A. J. Kearsley. Transport Phenomena in Biological Field Effect Transistors. *SIAM Journal on Applied Mathematics* **80**:6 (2020), 2586–2607. DOI: 10.1137/19M1255495

[4] R. M. Evans, A. Balijepalli, and A. J. Kearsley. Diffusion-limited Reactions in Nanoscale Electronics. *Methods and Applics. of Analysis* **26**:2 (2019), 149–166.

[5] P. Flajolet, X. Gourdon, and P. Dumas. Mellin Transforms and Asymptotics: Harmonic Sums. *Theoretical Computer Science* **144**:1 (1995), 3–58.

[6] A. J. Kearsley, Y. Gadhyan, and W. E. Wallace. Stochastic Regression Modeling of Chemical Spectra. *Chemometrics and Intelligent Laboratory Systems* **139** (2014), 26–32.

Quantifying Flows in Time-Irreversible Markov Chains: Application to Budding Yeast Gene Regulatory Network

Danielle Middlebrooks
Geoffrey McFadden
Maria Cameron (University of Maryland)

Gene Regulatory Networks (GRNs) are a powerful modeling tool for cellular processes that can reduce the need to estimate the large number of parameters required for a general kinetic model. GRNs consist of (i) nodes, each of which may be in one of two states, active or passive, (ii) two types of directed edges representing activation or deactivation, and (iii) a set of transition rules. Any GRN with N nodes has a one-to-one mapping to a discrete-time Markov chain with 2^N states, which is typically time-irreversible. In order to quantify the dynamical evolution of the system, we develop a methodology within the framework of Transition Path Theory (TPT) for the efficient description of the transition process from a starting state to a target state. TPT is a powerful tool for the analysis of transitions in Markov chains between two disjoint subsets of states and can be used to analyze efficiently both discrete-time and continuous-time Markov chains originating from GRNs.

We have applied the proposed methodology to a budding yeast cell cycle GRN with two types of transition rules, deterministic or stochastic, and have demonstrated that the stochastic model is notably more robust with respect to edge removals. Our analysis suggests that random factors provide compensatory mechanisms for the cell cycle to proceed even if one of the essential proteins/protein complexes is dysfunctional. A Markov chain originating from a GRN is typically irreversible, and the graph it induces is cyclic. We replaced the effective current along each edge by an acyclic current and offered an algorithm for calculating it. The acyclic current induces an acyclic graph whose trajectories start in one subset, A, of the network and run until reaching a sink that necessarily belongs to a disjoint subset, B.

Recent efforts have focused on improving the computational efficiency of the original algorithm. The applicability of TPT to the full Markov chain corresponding to a GRN with N nodes is typically memory-limited to $N \leq 20$ as the Markov chain has 2^N states. Through our analysis, we have observed that even though the dynamical network for some GRNs become huge, there are numerous states which have extremely low probabilities of being reached by the biological process being modeled. Most recently, we incorporated in our algorithm a way to extract the most meaningful states, by using Monte Carlo sampling to apply our

methodology without the need to store the entire network in computer memory.

A future goal is to use our methodology to answer questions arising in more complicated GRNs. One network of interest is the GRN model representing so-called segment polarity genes. This group of genes is involved in embryonic pattern formation in the fruit fly Drosophila Melanogaster. Homologs of the segment polarity genes have also been identified in vertebrates, including humans, which suggests strong evolutionary conservation. This GRN is more computationally challenging in that these genes refine and maintain their expression through a complex network of intracellular and intercellular regulatory interactions.

Mathematically Investigating Retinitis Pigmentosa

Danielle Brager
Erika T. Camacho (Arizona State University)
Thierry Léveillard (Institut de la Vision, France)
Stephen Wirkus (Arizona State University)

Retinitis Pigmentosa (RP), the most common inherited retinal degeneration, is a collection of clinically and genetically heterogeneous degenerative retinal diseases. There is no cure for RP. Patients with RP experience progressive irreversible vision loss due to the sequential degeneration of rod and cone photoreceptors, the light-sensing cells in the retina that play an essential role in the vision process. While known genetic mutations associated with RP affect the rods, triggering their dysfunction and death, the loss of cones inevitably follows in a manner independent of those genetic mutations. Photoreceptor degeneration is the hallmark of RP and is characterized by the shortening of the photoreceptors' outer segments (OS) which is the subcellular structure that carries the photoreceptor's light sensitive molecules. Still, the secondary wave of cone photoreceptor death is an anomaly, and investigation of this phenomenon resulted in the discovery of the rod-derived cone viability factor (RdCVF). RdCVF is a protein secreted by the rods that preserves the cones by accelerating the flow of glucose into the cone cells stimulating aerobic glycolysis. Cones are responsible for color vision as well as visual acuity and rely on aerobic glycolysis to supply the metabolites necessary for OS renewal and maintenance.

We have formulated a predator-prey style system of nonlinear ordinary differential equations (ODEs) to mathematically investigate photoreceptor interactions in the presence of RP while accounting for the new understanding of RdCVF's role in enhancing cone survival [1]. We model photoreceptor population dynamics in the context of OS lengths treating the renewal and shedding of OS as a birth and death process. We are investigating RP once it has already started to degenerate the rods. Thus, we separate the rod population according to phenotype, normal or mutated. We assume that there is a shared nutrient source for all photoreceptors which is mediated by the neighboring retinal pigment epithelium (RPE) and call this nutrient source the trophic pool. We used the mathematical model and subsequent analysis to examine the underlying processes and mechanisms (defined by the model parameters) leading to blindness in RP patients.

The analysis of this model included a local stability analysis, bifurcation analysis, and a global sensitivity analysis. In all our analyses, both analytical and numerical, a set of key parameters involved in OS shedding, OS renewal, and nutrient supply were shown to govern the dynamics of the system. The members of this set of parameters formed five ratios that explicitly appeared in coordinates of the model equilibrium points as well as conditions necessary for the physiological relevance, stability, and coalescence of equilibrium points. The five key ratios are: OS shedding to OS renewal of normal rods, OS shedding to OS renewal of mutated rods, OS shedding to OS renewal of cones with RDCVF, OS shedding to OS renewal of cones without RdCVF, and the nutrient carrying capacity

The mathematical model has nine equilibrium points with only seven of them being physiologically relevant. We interpreted the seven physiologically relevant equilibrium points as different stages of retinal degeneration in which some combination of normal rods, mutated rods, cones, or nutrients are gone. We determined conditions necessary for the local asymptotic stability of the equilibrium points. Mathematically, these conditions present as inequalities composed of model parameters. Therefore, we used the results as criteria needed to remain in a stage in the progression of retinal degeneration. We found that our mathematical model captures various stages of RP, but also captures the progression seen in cone-rod dystrophy (CRD). CRD is a group of inherited retinal diseases where the loss of cones precedes the loss of rods. Following the stability analysis, a bifurcation analysis was performed to explore pathways to blindness. The bifurcation analysis revealed five pathways to blindness where two of those paths illustrate progressions of RP and the other three show progressions of CRD. Both the stability and bifurcation analysis results reveal that the ratio of OS shedding to renewal and the availability of nutrients may play a key role in halting RP progression.

Numerical analyses including a global sensitivity analysis and a numerical bifurcation analysis were performed after experimental data was used for parameter estimation and model validation. Four of the model equilibrium points were present in the two RP pathways to blindness found during the bifurcation analysis previously discussed. Therefore, we focused on those four

equilibrium points when conducting our numerical analyses. The numerical results are framed in terms of stages in the progression of RP. Using Latin Hypercube Sampling coupled with partial correlation coefficients, we performed a sensitivity analysis to determine mechanisms that have a significant effect on the cones at different stages of RP. Once again, these results highlight the importance of the delicate balance between OS shedding and renewal. Finally, we derived a non-dimensional form of the mathematical model and performed a numerical bifurcation analysis via MATCONT to explore the existence of stable limit cycles. Using MATCONT, we searched for Hopf bifurcations and found that stable limit cycles are born from two of the equilibrium points involved in the RP blindness pathway. A stable limit cycle is a stable mode, other than an equilibrium point, where the rods and cones coexist. Thus, we take the presence of a stable limit cycle to be an indication of a phase where therapy and prevention of the progression of RP may occur.

Our findings illustrate the benefit of using mathematical models to uncover mechanisms driving the progression of RP and opens the possibility to use in silico experiments to test treatment options in the absence of rods. Future work on this problem could include an analytical Hopf bifurcation analysis, optimal control of the model, and analysis of an extended model that incorporates molecular level interactions involved in the biochemical cascade triggered by glucose uptake.

[1] D. C. Brager. *Modeling and Analyzing the Progression of Retinitis Pigmentosa*. Doctoral dissertation, Arizona State University, 2020.

Materials Modeling

Mathematical modeling, computational simulation, and data analytics are key enablers of emerging manufacturing technologies. The Materials Genome Initiative (MGI), an interagency program with the goal of significantly reducing the time from discovery to commercial deployment of new materials using modeling, simulation, and informatics, is a case in point. To support the NIST role in the MGI, we develop and assesses modeling and simulation techniques and tools, with emphasis on uncertainty quantification, and collaborate with other NIST Laboratories in their efforts to develop the measurement science infrastructure needed by the materials science and engineering community.

OOF: Finite Element Analysis of Material Microstructures

Stephen A. Langer
Günay Doğan
Andrew C.E. Reid (NIST MML)
Shahriyar Keshavarz (Theiss Research)

http://www.ctcms.nist.gov/oof/

The OOF Project, a collaboration between ACMD and MML, is developing software tools for analyzing real material microstructure. The microstructure of a material is the (usually) complex ensemble of polycrystalline grains, second phases, cracks, pores, and other features occurring on length scales large compared to atomic sizes. The goal of OOF is to use data from a micrograph of a real or simulated material to compute its macroscopic behavior via finite element analysis.

The OOF user loads images into the program, assigns material properties to the features of the image, generates a finite element mesh that matches the geometry of the features, chooses which physical properties to solve for, and performs virtual experiments to determine the effect of the microstructural geometry on the material. OOF is intended to be a general tool, applicable to a wide variety of microstructures in a wide variety of physical situations. OOF2 and OOF3D are used by educators and researchers in industry, academia, and government labs worldwide.

There are two versions of OOF, OOF2 and OOF3D, each freely available on the OOF website. OOF2 starts with two dimensional images of microstructures and solves associated two-dimensional differential equations, assuming that the material being simulated is either a thin freely suspended film or a slice from a larger volume that is unvarying in the third dimension (generalizations of plane stress and plane strain, respectively). OOF3D starts with three dimensional images and solves equations in three dimensions. Development this year continued on multiple fronts.

A plastic deformation, as opposed to an elastic deformation, is one in which a deformed material does not return to its original conformation after applied stresses are removed. Crystal plasticity describes the deformation of crystalline materials, in which the behavior is affected by the conformation of the atomic lattice. Computational models of plastic deformation are complex, non-linear, and (sometimes) controversial. A. Reid and S. Keshavarz have been working to implement crystal plasticity in OOF3D for both single crystals and polycrystals (assemblies of differently oriented grains). Although the current implementation is specialized to FCC crystal structures and a power-law constitutive rule (the relationship between stress and strain or strain rate), care has been taken to ensure that the software infrastructure is sufficiently generic that different crystal structures and different constitutive rules can be implemented without architectural changes, preserving the generality and extensibility that is the hallmark of the OOF approach.

Future work will include large-strain elasticity, additional types of plasticity, and investigating the use of machine learning to short-circuit the computationally expensive parts of the nonlinear computation.

G. Doğan has been developing algorithms to automate segmentation and meshing of microstructure images. Segmentation identifies distinct regions in a microstructure, and a mesh properly aligned with the segmented regions and their boundaries leads to accurate finite element simulations of the microstructure physics. In FY 2020, Doğan worked on two different methods for image segmentation. The first method builds on intrinsic representations of image regions embedded into phase field functions. Initial guesses of such phase field functions are evolved iteratively to detect the correct regions of microstructures, and to delineate them with well-defined boundaries. A 2d implementation of this method is currently available, and the extension to 3d is underway. The second algorithm, based on topology optimization, assigns region labels to image pixels in an iterative manner. This works well for multiphase segmentation. It can also handle more general statistics of the pixels within the regions, in contrast with piecewise constancy assumption of the first algorithm.

Doğan has been working with Prashant Athavale of Clarkson University on algorithms for restoration of electron backscattering diffraction (EBSD) images. EBSD is an important imaging modality to analyze and understand microstructure images by measuring the

crystal orientations at individual locations and generating a map of the orientations. However, EBSD maps are often noisy and incomplete; orientation measurements are missing in some locations, and some measurements contain errors. Doğan and Athavale implemented a reconstruction algorithm using a weighted total variation equation, a nonlinear partial differential equation that applies regularizing diffusion selectively to pixels. In this way, they were able to obtain state-of-the-art restorations of EBSD images. More details about this work on image segmentation and restoration can be found in the section of this report on Computational Tools for Image and Shape Analysis (see page 37).

S. Langer and A. Reid have been working with A. Creuziger of MML who is using OOF2 to study the effects of texture (crystal orientation distribution) on the properties of rolled steel. This has involved solving new types of equations, modifying the program to produce new outputs, adding EBSD based image segmentation methods (primitive versions of the ones being developed by Doğan), debugging, and optimization. OOF2 was able to reproduce an unexpected dip in experimentally measured stress curves and provided insight into the underlying physical mechanism.

OOF2 and OOF3D still rely on some old third-party software that will soon become obsolete. Langer has continued to work on the switch to the new versions of the aging software. Upgrading from gtk+2 to gtk3 (used for the graphical user interface) needs to be done before the switch from Python 2 to Python 3 (used for high level programming and scripting) because there is no support for gtk+2 in Python 3.

Converting OOF2 to gtk3 is nearly complete. The bulk of the code has been converted, and a new graphics canvas has been written, because there is no gtk3 version of the external canvas library that was used with gtk+2. OOF2 uses a testing tool developed for the gtk+2 version that allows the developers to record a session, annotate the resulting log file with tests, and replay the annotated log file. The log files form a test suite that ensures that the GUI is working correctly. This tool, gtklogger, has been rewritten for gtk3, and the test suite for OOF2 is being assembled. The new canvas code and gtklogger will both be useful outside of OOF and will be released separately.

OOF2 and OOF3D continue to be used outside of NIST. OOF2 was downloaded about 700 times this past year, and OOF3D was downloaded about 200 times. OOF2 can be run on the NSF nanoHUB facility,[17] where it was used in approximately 4000 sessions in 2020. It has been used by more than 3000 people on nanoHUB since it was installed in late 2007.

Micromagnetic Modeling

Michael Donahue
Donald Porter
Robert McMichael (NIST PML)
Cindi Dennis (NIST MML)

http://math.nist.gov/oommf/

Advances in magnetic devices such as recording heads, field sensors, spin torque oscillators, and magnetic nonvolatile memory (MRAM) are dependent on an understanding of magnetization processes in magnetic materials at the nanometer level. Micromagnetics, a mathematical model used to simulate magnetic behavior, is needed to interpret measurements at this scale. ACMD is working with industrial and academic partners, as well as with colleagues in MML and PML to improve the state-of-the-art in micromagnetic modeling.

We have developed a public domain computer code for performing computational micromagnetics, the Object-Oriented Micromagnetic Modeling Framework (OOMMF). OOMMF serves as an open, well-documented environment in which algorithms can be evaluated on benchmark problems. OOMMF has a modular structure that allows independent developers to contribute extensions that add to its basic functionality. OOMMF also serves as a fully functional micromagnetic modeling system, handling three-dimensional problems, with extensible input and output mechanisms. In FY 2020 alone, the software was downloaded more than 5700 times by more than 3700 distinct client machines. The use of OOMMF was acknowledged in more than 210 peer-reviewed journal articles. OOMMF has become an invaluable tool for the magnetics research community.

Developments in the last year include the following.

- Presented a four-session (eight-hour total) tutorial series, Micromagnetics and OOMMF, in May and June 2020 as part of the Online Spintronics Seminar sponsored by the IEEE Magnetics Society.[18]
- Answered approximately 125 questions in the post-tutorial Q&A forums.
- Increased support for periodic boundary conditions in OOMMF.
- Released version 1.2b4 of OOMMF, incorporating updates to an extension supporting GPUs.[19]

OOMMF is part of a larger activity, the Micromagnetic Modeling Activity Group (muMAG), formed to address fundamental issues in micromagnetic modeling through two activities: the development of public domain reference software, and the definition and

[17] https://nanohub.org/tools/oof2

[18] https://www.spintalks.org/tutorials

[19] https://math.nist.gov/oommf/software-12.html

dissemination of standard problems for testing modeling software. ACMD staff members are involved in development of the standard problem suite as well. The history of muMAG standard problems is the subject of a book chapter published this year [1].

In addition to the continuing development of OOMMF, the project also does collaborative research using OOMMF. One such collaboration, with researchers at the University of Akron, involves using micromagnetic simulations to improve performance in electric motors [2][3]. M. Donahue is a team member on the NIST Innovations in Measurement Science nanothermometry project Thermal MagIC,[20] and he also provides technical guidance on micromagnetic modeling for the DARPA M3IC (Magnetic, Miniaturized, and Monolithically Integrated Components) project,[21] which aims to integrate magnetic components into the semiconductor materials fabrication process to improve electromagnetic systems for communications, radar, and related applications.

In total, during this period the ACMD micromagnetic project produced a book chapter [1], two journal articles [2][3], and four conference presentations [4][5][6][7]. Notably, M. Donahue received the 2020 Award for Excellence in Research in Applied Mathematics from the Washington Academy of Sciences in recognition of "...excellence in applied mathematics, leading to new tools for modeling and simulation which have transformed research into nanoscale magnetic films, structures and devices." [22]

[1] D. G. Porter and M. J. Donahue. Standard Problems in Micromagnetics. In *Electrostatic and Magnetic Phenomena* (M. J. Donahue, ed.), World Scientific, 2020, 285–324. DOI: 10.1142/9789813270268_0009

[2] F. Ahmadi, Y. Sozer, M. J. Donahue, and I. Tsukerman. A Low-loss and Lightweight Magnetic Material for Electrical Machinery. *IET Electric Power Applications* **14** (2020), 282–290. DOI: 10.1049/iet-epa.2019.0430

[3] F. Ahmadi, M. J. Donahue, Y. Sozer, and I. Tsukerman. Micromagnetic Study of Magnetic Nanowires. *AIP Advances* **9** (2019), 125047. DOI: 10.1063/1.5130157

[4] A. J. Biacchi, E. De Lima Correa, F. Zhang, T. Moffat, W. Tew, M. J. Donahue, S. Woods, C. Dennis, and A. R. Hight Walker. "Tuning the Properties of Colloidal Magnetic Particles for Thermometry on the Nanoscale." 64th Annual Conference on Magnetism and Magnetic Materials (MMM 2019), Las Vegas, NV, November 8, 2019.

[5] F. Ahmadi, M. J. Donahue, Y. Sozer, and I. Tsukerman. "Micromagnetic Study of Soft Magnetic Nanowires." 64th Annual Conference on Magnetism and Magnetic Materials, Las Vegas, NV, November 7, 2019.

[6] M. J. Donahue and D. G. Porter. "Quantitative Evaluation and Reduction of Error in Computation of the Demagnetization Tensor." 64th Annual Conference on Magnetism and Magnetic Materials, Las Vegas, NV, November 7, 2019.

[7] M. J. Donahue and D. G. Porter. "High Order Methods for Computing the Demagnetization Tensor for Periodic Boundaries." 65th Conference on Magnetism and Magnetic Materials, Online, November 6, 2020.

Resolving the Shock Layer in Adsorption Breakthrough Measurements

Robert DeJaco
Anthony Kearsley
Paul Patrone

Separation processes are essential for many of our basic needs like the air we breathe, the water we drink, and the medications we take. In an adsorption process, a cylindrical column is packed with a solid material and mixture is fed into one end of the column. As the fluid flows, the solute adsorbs preferentially over the solvent and takes a longer time to exit the column than the solvent, yielding a separation that can be captured at the outlet. Adsorption processes are typically more energetically efficient than conventional techniques that rely on differences in boiling, as they instead rely on differences in affinities for a solid. Such processes have the potential to be more advantageous as the pool of candidate solid materials increases at an increasing rate.

When a new adsorbent material with superior separation potential of a given mixture is first conceptualized, a column is packed with the solid and the mixture to be separated is fed into the inlet. A measurement of the concentration exiting, or breaking through, the column, as a function of time is referred to as the break-through curve. Since the portion of the mixture that adsorbs preferentially takes longer to exit the column, the extent of separation can be quantified by the break-through curve measurement. As the extent of separation observed in the measurement determines whether time and money should be invested in the assessment of the process in larger industrial scales, understanding the dynamic measurement is therefore essential to efficient industrial implementation of adsorption processes.

[20] https://www.nist.gov/programs-projects/thermal-magic-si-traceable-method-3d-thermal-magnetic-imaging-and-control

[21] https://www.darpa.mil/program/magnetic-miniaturized-and-monolithically-integrated-components

[22] https://www.washacadsci.org/awards-ceremony-2020/

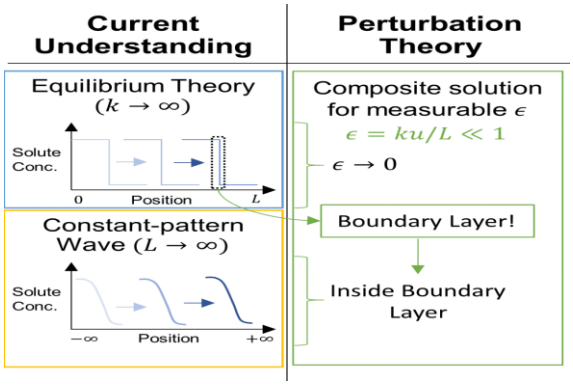


Figure 47. *Our perturbation theory approach provides a holistic understanding of two classical limiting interpretations of the physical phenomena arising in adsorption dynamics. It demonstrates how such phenomena are connected and can be observed in real experiments where mass transfer and column length cannot be infinite, and that they really occur in the limit of slow convection relative to adsorption, $\epsilon = ku/L \ll 1$.*

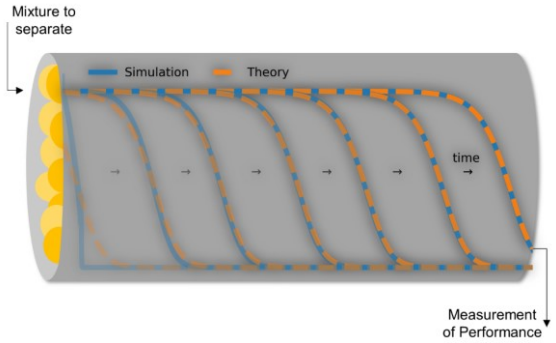


Figure 48. *Comparison between numerical simulations (blue) and perturbation theory (orange) in the prediction of fixed-bed adsorption dynamics, overlaid upon a sketch of solid adsorbent particles (yellow) packed in an adsorption column (grey). These calculations are performed for Langmuir adsorption isotherms with $\epsilon = 0.04$. The progression of time is depicted with decreasing transparency.*

As illustrated in Figure 47, the understanding of the dynamics of fixed-bed adsorption [1] is primarily confined to two limiting scenarios: the first being the case of infinitely-fast adsorption ($k \to \infty$, often referred to as "equilibrium theory"), and the second being the presence of a constant-pattern wave for very long columns ($L \to \infty$). Using perturbation theory in the case of single-solute adsorption with plug flow in the limit of a fast rate of adsorption relative to rate of convection, $\epsilon = kuL/ \ll 1$, we have found that these two phenomena are related. Equilibrium theory corresponds to the leading order problem when $\epsilon \to 0$. Shock waves (discontinuities) associated with equilibrium theory correspond to a boundary layer. Inside this boundary layer is the constant pattern wave that has previously only been shown to exist for columns of infinite length. Our asymptotic approach connects these concepts and demonstrates when the phenomena can occur for finite, measurable systems where $\epsilon \ll 1$ (i.e., ϵ need not be 0).

A typical comparison between perturbation theory and numerical simulation is depicted in Figure 48, overlaid upon a cartoon of a fixed bed adsorption column. At short times, boundary layers are present at the column inlet moving with the carrier fluid. These boundary layers are $O(\epsilon)$ in space and time. The long-time composite solution (orange) of the leading order problems inside and outside of the boundary layer is constant with time with respect to a Galilean coordinate transformation. It agrees well with the numerical simulations after being shifted to account for the short-time boundary layer. Conveniently, the theoretical predictions are accurate for large times before which the solute exits the column. As a result, the theoretical predictions can be considered a good approximation of the shape of the break-through curve measurement.

Our work provides a unifying framework for understanding the measurement in more detail. Since equilibrium theory is often used in the initial design of complex cyclic adsorption processes involving multiple interacting beds [2], we anticipate that using perturbation theory to resolve the associated boundary layers will also result in more accurate designs. We are currently writing up these results for publication. In the future, we plan to extend our analysis to cases where thermal variations and axial dispersion are significant.

[1] R. T. Yang. *Gas Separation by Adsorption Processes*, Butterworth, 1987, 141-200.

[2] M. Mazzotti and A. Rajendran. Equilibrium Theory-Based Analysis of Nonlinear Waves in Separation Processes. *Annual Review of Chemical and Biomolecular Engineering* **4** (2013), 119-141.

A Kinetic Model of Foam Wall Ruptures

Joe Klobusicky

In the field of mathematical materials science, rigorous treatment of the coarsening of foam due to wall ruptures are much less studied than coarsening through gas diffusion. This is likely due to several complicating factors involving the coarsening process. First, unlike gas diffusion in which cells continuously shrink to zero area, edge rupture is a relatively violent and discontinuous event, and can create intricate configurations of cells which are quite difficult to properly catalogue and model. Another issue concerns the stabilization of networks immediately following a wall rupture, in which cell edges reconfigure to minimize surface tension, while maintaining the network's topology. Calculating this quasi-stationary state of the network boundary is complicated and is dependent on the network's geometry. On the other hand, foam rupture has a much simpler

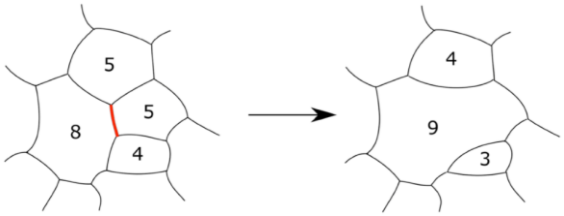


Figure 49. *Side numbers of cells neighboring a wall (shown in bold red) before and after rupture.*

"natural clock" which is simply the total number of ruptures. In gas diffusion, topological transitions occur when cells shrink to zero area, which corresponds to the occurrence of singularities in curvature flow, and thus are not known a priori.

In this project, we are interested in constructing a minimal model for foam breaking. As such we do not consider geometric quantities such as curvature and cell areas, but rather only keep track of each cell's topology, or its number of sides. In most instances, when a wall is ruptured, topology rearrangement can be viewed as two simultaneous merging events (see Figure 49):

- *Face merging:* Cells which share a face with the ruptured wall join together. If the two cells have i and j sides before rupture, afterward the new cell has $i + j - 4$ sides.
- *Edge merging:* Cells which share a single vertex with the ruptured wall will each lose a side.

For a cell C_n containing n sides, a standard wall rupture consists of one face merging event, and two edge merging events. The reaction can be represented as

$$C_i + C_j \to C_{i+j-4}, \quad C_k \to C_{k+1}, \quad C_l \to C_{l-1}.$$

By imposing probabilities on which cell walls rupture, we can iterate the foam braking process, which is in fact a Markov process on the space of trivalent graphs. Figure 50 shows snapshots of a sample path.

Since face merging is a second order collision (two particles merge into one), it suggests that the fundamental dynamics for foam coarsening may be related to coagulation processes. In fact, the collision rate for these equations is similar to that of the Smoluchowski coagulation equations with multiplicative kernel, which models cluster sizes for a system of sticky particles. Such equations are known to have gelation effects, in which a single particle of infinite size occurs [2]. Under an appropriate mean field assumption for how neighbors are selected when a wall ruptures, we may write limiting kinetic equations for the proportion $u(n,t)$ of n-sided cells at time t. These form an infinite system of birth death equations, with

$$\dot{u}(n,t) = \begin{matrix} Q_+^F(n,t) + Q_+^E(n,t) \\ -Q_-^F(n,t) - Q_-^E(n,t) \end{matrix}, \quad n \geq 3.$$

The four terms $Q_{+/-}^{F/E}(n,t)$ on the right-hand side generally involve products and quotients of $u(k,t)$ for various $k \geq 3$, which makes this system nonlinear. We can simulate solutions to this system by an easy to implement particle system method similar that used in [1] for grain coarsening by curvature flow. We hope to provide in future works a wide collection of experiments in the future which compare solutions to kinetic solutions to statistics of the Markov process on networks.

[1] J. Klobusicky, G. Menon, and R. L. Pego. Two-Dimensional Grain Boundary Networks: Stochastic Particle Models and Kinetic Limits. arXiv preprint arXiv:1810.07828.

[2] F. Leyvraz and H. R. Tschudi. Singularities in the Kinetics of Coagulation Processes. *Journal of Physics A: Mathematical and General* **14**:12 (1981), 3389.

[3] J. Klobusicky. Markov Models of Coarsening in Two-Dimensional Foams with Edge Rupture. arXiv preprint arXiv:2101.0096.

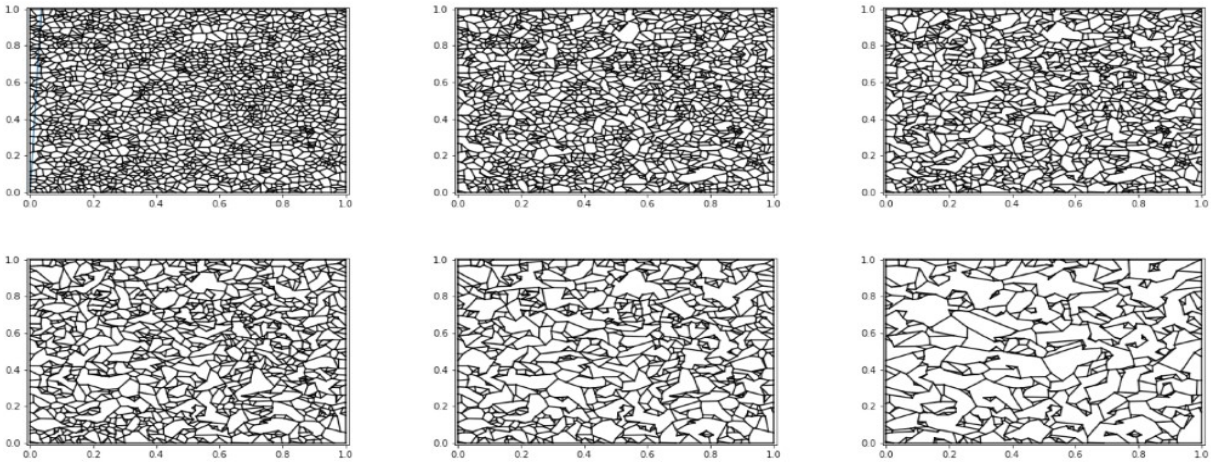


Figure 50. *Snapshots of the wall rupturing process. Top row: Network of 2000 initial cells after no ruptures (left), 300 ruptures (middle), and 600 ruptures (right). Bottom row: The network after 900 ruptures, 1200 ruptures, and 1500 ruptures.*

Coarsening of Voronoi Diagrams

Joe Klobusicky
Anthony Kearsley
Jason Evelyth (Brown University)

Voronoi diagrams are important tools in computational geometry, applicable to a variety of subjects including materials science, economics, and computer graphics [1]. For our application in materials science, we are interested in Voronoi diagrams generated by n distinct site points on a square domain. For each site point p, the corresponding Voronoi cell V_p is defined as points in the domain which are closer to p than any other site point. It can be shown that each Voronoi cell is a convex polygon, and that the collection of Voronoi cells forms a tiling of the domain.

Voronoi diagrams offer a novel way to consider the coarsening of trivalent networks, in which all vertices incident with three edges. This done by evolving site points to minimize a prescribed energy functional defined on the network. The purpose of this project is to construct efficient algorithms which perform Voronoi diagram coarsening over a wide range of functionals, and to compare them with more popular coarsening methods, such as coarsening by curvature flow.

A major component for performing efficient coarsening of Voronoi diagrams is building a sufficiently fast algorithm which is amenable to variations in the network's domain (e.g., periodic boundaries or the unit square). This was achieved with a careful implementation of Fortune's algorithm, a sweep-line algorithm which can construct Voronoi diagrams with a time complexity of $O(n \log n)$, as opposed to brute force methods which take $O(n^2)$ operations.

One attractive feature of Voronoi coarsening is that minimization is naturally performed over a space with dimension equal to the number of Voronoi cell site points. This should be contrasted with curvature-based coarsening in which points evolve with respect to a complicated system of parabolic partial differential equations. As coarsening evolves, an interesting variety of collisions can occur between site points. Some collisions require the removal of site points (and therefore Voronoi cells) to continue the flow. This should be seen as a coarsening step, as the removal of cells increases average cell area and induces a coarsening rate for the average cell area increase over time.

Implementation of the minimization process can be obtained either through a quasi-Newton method or stochastic gradient descent. We note that a rigorous investigation for coarsening driven by an energy function equal to the total perimeter, including bounds on coarsening rates, was undertaken by Elsey and Slepcev [2]. See Figure 51 for a visualization of the process.

In principle, we can consider minimization over a wide variety of energy functionals. For example, through a penalization term we may consider an energy which induces a repulsion between site points. For each such energy, we can compare network statistics of the coarsened network against other well-known coarsening processes. We hope to investigate a wide range of these functionals in the future.

[1] M. De Berg, M. Van Kreveld, M. Overmars, and O. Schwarzkopf. Computational Geometry. In *Computational Geometry*, Springer, Berlin, 1997 1-17.

[2] M. Elsey and D. Slepčev, D. (2015). Mean-curvature Flow of Voronoi Diagrams. *Journal of Nonlinear Science* **25**:1 (2015), 59-85.

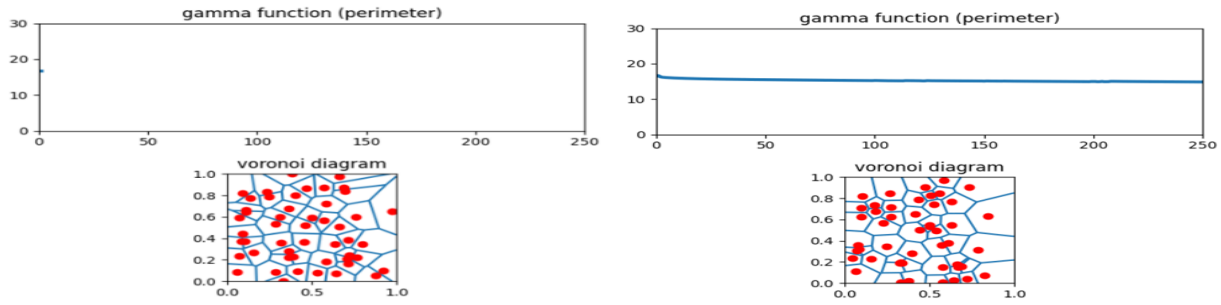


Figure 51. *A visual of the Voronoi coarsening process driven by minimization of total perimeter. Left: Initial network of 50 cells. Right: The network after Voronoi coarsening through minimization of perimeter. Note the minimization process makes site points tend to cluster.*

Extending Zeno

John Nolan
Jack Douglas (NIST MML)
Debra Audus (NIST MML)

https://zeno.nist.gov/

Work on generalized random walks based on multivariate stable laws has been extended. The primary motivation for this is to expand the capabilities of the Zeno program, a NIST package to probe the structure of complex objects, e.g. polymers. Other possible uses are for simulating solutions to fractional diffusion equations, which are used to model fluid movement in porous materials.

Our focus has been on improvements to the algorithms developed last year. Speeding up the simulation of hitting distributions for stable processes has been a major focus, because Zeno-type simulations involve millions of iterations. Also, the initial focus was on dimension $d = 3$, but the results are now extended to any dimension. The Walk-On-Spheres method is a method to accelerate simulations in the Brownian case. An extension, called Walk-In-And-Out-Of-Balls, has been developed that considers the discontinuous paths of stable motion.

[1] J. P. Nolan. *Univariate Stable Distributions*. Springer Verlag, 2020, xv + 333 pages.

High Performance Computing and Visualization

Computational capability continues to advance rapidly, enabling modeling and simulation to be done with greatly increased fidelity. Doing so often requires computing resources well beyond what is available on the desktop. Developing software that makes effective use of such high-performance computing platforms remains very challenging, requiring expertise that application scientists rarely have. We maintain such expertise for application to NIST problems. Such computations, as well as modern experiments, typically produce large volumes of data, which cannot be readily comprehended. We are developing the infrastructure necessary for advanced interactive, quantitative visualization and analysis of scientific data, including the use of 3D immersive environments, and applying the resulting tools to NIST problems

High Precision Calculations of Fundamental Properties of Few-Electron Atomic Systems

James Sims
Maria Ruiz (University of Erlangen, Germany)
Bholanath Padhy (Khallikote College, India)

NIST has long been involved in supplying critically evaluated data on atomic and molecular properties such as the atomic properties of the elements contained in the Periodic Table and vibrational and electronic energy level data for neutral and ionic molecules contained in the NIST Chemistry WebBook. Fundamental to this work is the ability to predict, theoretically, a property more precisely than even the most precise experiments. It is our goal to be able to accomplish this for few-electron atomic systems.

While impressive advances have been made over the years in the study of atomic structure in both experiment and theory, the scarcity of information on atomic energy levels is acute, especially for highly ionized atoms. The availability of high precision results tails off as the state of ionization increases, not to mention higher angular momentum states. In addition, atomic anions have more diffuse electronic distributions, representing more challenging computational targets than the corresponding ground states.

In the past two decades, there has been breathtaking improvements in computer hardware and innovations in mathematical formulations and algorithms, leading to "virtual experiments" becoming a more and more cost-effective and reliable way to investigate chemical and physical phenomena. Our contribution in this arena has been undertaking the theoretical development of our hybrid Hylleraas-CI (Hy-CI) wave function method to bring sub-chemical precision to atomic systems with more than two electrons.

Hy-CI has from its inception been an attempt to extend the success of the Hylleraas (Hy) method to systems with more than three electrons, and hence is an attempt to solve not just the three-body problem but the more general N-body problem [1]. Fundamental to the method is the restriction of one r_{ij} per configuration state function (CSF). (For atomic systems with greater than four electrons, all relatively precise calculations nowadays adopt the Hy-CI methodology of one r_{ij} term per CSF). In the case of three electron lithium systems, we have computed four excited states of the lithium atom to two orders of magnitude greater than has ever been done before [2]. At the four-electron level, to get truly precise chemical properties like familiar chemical electron affinities and ionization energies, it is important to get close to the nanohartree level we achieved for the three-electron atom, a significantly more difficult problem for four electrons than for three. By investigating more flexible atomic orbital basis sets and better configuration state function filtering techniques to control expansion lengths, we have been able to successfully tackle the four-electron case.

Progress to date has included computing the non-relativistic ground state energy of not only beryllium, but also many members of its isoelectronic sequence to eight significant digit precision. With the results from our calculations and a least squares fit of the calculated energies, we have been able to compute the entire beryllium ground state isoelectronic sequence for Z = 4 through Z = 113 [3]. Li^- (with Z=3), nominally the first member of this series, has a decidedly different electronic structure and was not included in those calculations and subsequent discussions, but that omission has been corrected and we have subsequently carried out large, comparable calculation for the Li^- ground state [4].

The first member of the Be isoelectronic ground state sequence, the negative Li^- ion, is also a four- electron system in which correlation plays a very important part in the binding. However due to the reduced nuclear charge, it is a more diffuse system in which one of its outer two L shell electrons moves at a greater distance from the nucleus than the other, and hence its nodal structure is different from that of a coupled L shell with an identical pair of electrons. The ground state of the singlet S state of Li^- is the same type of problem as the first excited state of Be; it is like Be(2s3s), not Be(2s2s). Completing this calculation has provided the necessary insight to enable the calculation of the Be first excited

state of singlet S symmetry, Be(2s3s), to an order of magnitude better than previous calculations. Armed with this result, we have been able to continue this level of precision to the Be(2s4s) excited state and have calculated the higher, more diffuse Be(2s5s) through Be(2s7s) states as well this past year, and in the process have demonstrated that Hy-CI can calculate the higher, more diffuse Rydberg states with more complicated nodal structures to the same level of precision as the less excited states [5].

While our work has demonstrated the efficacy of Hy-CI as a solution to the N-body problem for four or more electrons, this work has also shown the presence of a "double cusp" $r_{12}r_{34}$ term type slow convergence problem at the nanohartree precision level which is ultimately built into Hy-CI for four or more electrons. We have investigated a generalization of the Hy-CI method to an exponentially correlated Hy-CI (E-Hy-CI) method in which the single r_{ij} of an Hy-CI wave function is generalized to a form which pairs an exponential r_{ij} factor with linear r_{ij}, producing a correlation factor which has the right behavior in the vicinity of the r_{ij} cusp, and also as r_{ij} goes to infinity. Our results for the ground singlet S state of the two-electron helium atom have demonstrated both convergence acceleration and an improvement in the precision for the same basis. This makes the application of the E-Hy-CI method to systems with N > 4, for which this formalism with at most a single exponentially correlated and linear r_{ij} factor leads to solvable integrals, very promising. The ground 1 S state non-relativistic energy of He is computed to be -2.9037 2437 7034 1195 9831 1084 hartrees (Ha) for the best expansion. We are currently calculating ground and excited S states of the Li^+ atom with the same technique, with results comparable to the He atom for the states computed so far.

[1] J. S. Sims and S. A. Hagstrom. Combined Configuration Interaction – Hylleraas Type Wave Function Study of the Ground State of the Beryllium Atom. *Physical Review A* **4**:3 (1971), 908. DOI: 10.1103/PhysRevA.4.908

[2] J. S. Sims and S. A. Hagstrom. Hy-CI Study of the 2^2S Ground State of Neutral Lithium and the First Five Excited ^2S States. *Physical Review A* **80** (2009), 052507. DOI: 10.1103/PhysRevA.80.052507

[3] J. S. Sims and S. A. Hagstrom. Hylleraas-Configuration Interaction Nonrelativistic Energies for the Singlet S Ground States of the Beryllium Isoelectronic Series up Through Z = 113. *Journal of Chemical Physics* **140** (2014), 224312. DOI: 10.1063/1.4881639

[4] J. S. Sims. Hylleraas-Configuration Interaction Study of the Singlet S Ground State of the Negative Li Ion. *Journal of Physics B: Atomic, Molecular and Optical Physics* **50** (2017), 245003. DOI: 10.1088/1361-6455/aa961e

[5] J. S. Sims. Hylleraas-Configuration Interaction (Hy-CI) Non-Relativistic Energies for the 3 Singlet S, 4 Singlet S, 5 Singlet S, 6 Singlet S and 7 Singlet S States of the Beryllium Atom. *Journal of Research of the National Institute of Standards and Technology* **125** (2020), 125006. DOI: 10.6028/jres.125.006

[6] J. S. Sims, B. Padhy and M. B. Ruiz. Exponentially Correlated Hylleraas-Configuration Interaction (E-Hy-CI) Non-Relativistic Energy of the 1 Singlet S Ground State of the Helium Atom. *International Journal of Quantum Chemistry* **121**:4 (2020), e26470. DOI: 10.1002/qua.26470

A Parallel Generalized Real Symmetric-Definite Eigenvalue Problem Solver

James Sims
Maria Ruiz (University of Erlangen, Germany)

Few-electron atomic and molecular systems continue to play an important role as a testbed for understanding fundamental physical and chemical properties. Non-relativistic ground state energies and wave functions for few-electron atoms are primary inputs for computations of atomic properties such as electron affinities and ionization potentials, as well as for computing perturbation corrections due to relativity and quantum electrodynamics. Furthermore, fundamental quantum theory dictates that, in general, uncertainties in wave functions scale as the square root of uncertainties in corresponding energies. Therefore, when using convergence of energies as a metric for convergence of the associated wave functions, very high precision in the former must be achieved to guarantee precision in the latter.

To achieve this high precision (real(24) arithmetic is necessary for two electron systems and real(16) for more than two electron systems) requires explicitly correlated wave functions which in turn leads to numerical linear dependency problems with sizes (N) that are large compared to most dense matrix problems. Such linear dependency issues are solved by going to higher precision. This, and the large system sizes, means that calculations need to be parallelized. Since parallel packages like ScaLAPACK are only available in real(8) they cannot be used as they stand, and they are too big to contemplate conversion to real(16).

This has led to the development of a portable parallel solver package, GRSDEP (generalized real symmetric-definite eigenvalue problem) for the generalized eigenvalue problem [1]. The solution technique used (shifted inverse iteration) is one of the fastest methods for getting selected roots of the generalized eigenvalue problem. The strategy for employing inverse iteration is described in detail in [1]. Also provided is a description of how this process is parallelized using MPI [2] for running the same program on multiple processors (on the same or different hosts). The timing tests which were done involved computing

the energy of a 4190-term hydrogen molecule wave function with a resultant total energy of -1.1744 7571 hartrees (Ha) [3] on differing numbers of processors. The 32 processor run speeds up the calculation by a factor of 20 over a single processor.

The parallel GRSDEP solver has proven quite useful, having been utilized for investigating (a) a 2-electron molecule: ground state of the hydrogen molecule [3], and (b) two-, three-, four-, and five-electron atoms and their ions, e.g. [4].

The software has been made publicly available via the NIST MIDAS repository,[23] where one finds the following description:

> Open source software for solving large-scale generalized eigenvalue problems on distributed computers. Suitable for large (80,000 by 80,000 or greater) dense matrices. Written in Fortran 90+. Includes a test program and sample output. In particular, this distribution, MPI_GEVP_ Package.tar.gz, consists of documentation (HowTo_MPI_GEVP_inviter.pdf), a collection of output files, and the software distribution itself. To use the software, download MPI_GEVP_ inviter_package.tar.gz, unwrap it, and follow the instructions in the HowTo to compile the solver and another program for generating test matrix elements. Then run various tests and compare tests with the output found in the various Output files.

[1] J. S. Sims and M. B. Ruiz. Parallel Generalized Real Symmetric-Definite Eigenvalue Problem. *Journal of Research of the National Institute of Standards and Technology* **125** (2020), 125032. DOI: 10.6028/jres.125.032

[2] Message Passage Interface Forum. MPI: A Message-Passing Interface Standard. *International Journal of Supercomputer Applications* **8**:3/4 (1994),159-416.

[3] J. S. Sims and S. A. Hagstrom. High Precision Variational Calculations for the Born-Oppenheimer Energies of the Ground State of the Hydrogen Molecule. *Journal of Chemical Physics* **124** (2006), 094191. DOI: 10.1063/1.2173250

[4] J. S. Sims and S. A. Hagstrom. Hylleraas-Configuration Interaction Nonrelativistic Energies for the Singlet S Ground States of the Beryllium Isoelectronic Series up Through Z = 113. *Journal of Chemical Physics* **140** (2014), 224312. DOI: 10.1063/1.4881639

[23] https://data.nist.gov/od/id/mds2-2293

Simulation of Dense Suspensions: Cementitious Materials

William George
Nicos Martys (NIST EL)
Clarissa Ferraris (NIST EL)
Steven Satterfield
Judith Terrill

A suspension is a collection of solid inclusions embedded in a fluid matrix. Suspensions play an important role in a wide variety of applications including paints, cement-based materials, slurries, and drilling fluids. Understanding the flow properties of a suspension is necessary in many such applications. However, measuring and predicting flow properties of suspensions remains challenging.

Suspensions can be quite complex, as the inclusions may have a wide range of shapes and a broad size distribution. Further complicating matters is that different matrix fluids may have quite disparate flow behavior. While the simplest type of matrix fluid is Newtonian, where the local stress is proportional to the shear rate, the matrix fluid can also be non-Newtonian, exhibiting quite complex behavior including shear thinning (viscosity decreases with shear rate), shear thickening (viscosity increases with shear rate), viscoelasticity (exhibiting both viscous and elastic properties), or even have a time dependent viscosity (thixotropic). We have two on-going studies on the rheology of cementitious materials, which are dense suspensions in non-Newtonian matrix fluids.

SRMs for Mortar and Concrete. Rotational rheometers, devices that measure fluid properties such as viscosity, are routinely used for homogeneous materials such as oils, but their use on dense suspensions, such as concrete, is a relatively new phenomenon. Since measurements with rheometers involve flow in a complex geometry, it is important that they are calibrated with a well characterized standard reference material (SRM). We are developing such SRMs in collaboration with the NIST Engineering Laboratory.

NIST produced an SRM for cement paste (SRM 2492) as the first step in the development of a reference material for concrete rheometers. The second step, the development of an SRM for mortar (SRM 2493) was completed in 2017 and is currently available. The material properties of the mortar SRM, such as viscosity, could not be measured in fundamental units with certainty. Thus, simulation was used to determine the viscosity of the mortar. To obtain the necessary fidelity in the simulations, a high-performance computing facility was used. Results of these simulations were compared with physical experiments as validation.

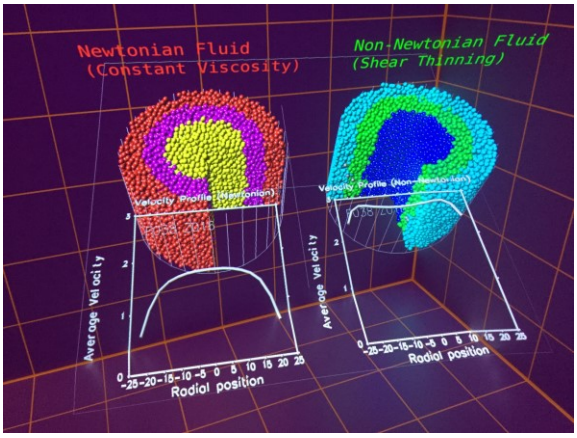


Figure 52. *Side-by-side comparison of identical systems with only a single property of the matrix fluid changed, one Newtonian and the other non-Newtonian (shear thinning).*

In 2019 we released SRM 2497, a standard reference concrete for rheological measurements [1]. The concrete SRM is comprised of the previously released mortar SRM with the addition of suspended 10 mm diameter hard glass spheres. The certified values for SRM 2497 were determined using the simulation results previously computed for the mortar SRM. As an indication of the impact of this work, the use of the cement paste and mortar SRMs are referenced in the new ASTM standard test method for measuring the rheological properties of cementitious materials [2].

We are currently running a suite of simulations of the concrete SRM using the cement paste as the matrix fluid in which both 1 mm and 10 mm diameter hard spheres are suspended. The results of these simulations should match the results from simulations comprising the mortar SRM as the matrix fluid with suspended 10 mm hard spheres. Depending on the outcome of these simulation, the concrete SRM certification may need to be updated. The current suite of concrete simulations is expected to complete in the next year.

Flow of Dense Suspensions in Pipes. Understanding the flow of suspensions in pipes is important for a wide variety of applications. For example, in the construction industry, concrete is often placed by pumping it through extensive pipe systems. However, research on predicting the pumpability of concrete has been limited due to the heavy equipment and large amounts of material needed. Suspension flow in pipes is also important in the developing field of 3D additive manufacturing.

Predicting the flow of this complex fluid, which is composed of a non-Newtonian matrix fluid with suspended solid inclusions flowing under pressure, is challenging. Flow in these systems is also complicated by variety of phenomena such as slip at the wall and shear induced migration, which has only been studied for the simpler case of a suspensions with a Newtonian matrix fluid. A detailed discussion of this topic is available in a NIST Technical Note [3]. It is also the case, especially in the case of 3D additive manufacturing, that the placement of these materials is time sensitive, from the time the material is initially mixed to the time it is pumped and placed [4].

We have been conducting detailed simulations of the flow of suspensions through pipes to enable the development of predictive flow models and to advance measurement science in this field. Through quantitative analysis and visualization of results, we have gained insight into shear-induced migration and slip behavior in these systems. For example, Figure 52 shows a side-by-side comparison of identical systems with only a single property of the matrix fluid changed, one Newtonian and the other non-Newtonian (shear thinning). Over the last year we have been conducting a suite of simulations of shear thinning and shear thickening suspensions, varying the properties of the matrix fluid and the driving force.

Studying the flow velocity fields as a function of driving force we have discovered a useful scaling relationship. Given that the matrix fluids have a viscosity, μ, that relates to the strain rate $\dot{\gamma}$ such that

$$\mu \sim \dot{\gamma}^n,$$

we have determined that the system velocity in the pipe, v, is related to the driving force, g, as

$$v \sim g^{\frac{1}{1+n}}$$

So, for example, with a shear thinning matrix fluid with $n = -0.5$, we have

$$v \sim g^{\frac{1}{1-0.5}} = g^2$$

and with a matrix fluid which is shear thickening, with $n = 0.5$, we have

$$v \sim g^{\frac{1}{1+0.5}} = g^{\frac{2}{3}}$$

Notice that this scaling relation depends on the power-law behavior of the non-Newtonian matrix fluid. As a consequence, given a few measurements of the flow velocity versus the driving force of the suspension, one can determine the power law behavior of the suspension, and indeed we can then also determine the power law behavior of the matrix fluid.

In the last year, all of our pipe-flow simulations began to reach steady state, although the simulation of shear thickening mortars in pipes has proven to be the slowest to approach steady state and so we will continue to run those simulations. Our results, which show agreement with the model predictions to within 15 %, were published in 2020 in the *Journal of Rheology* [5].

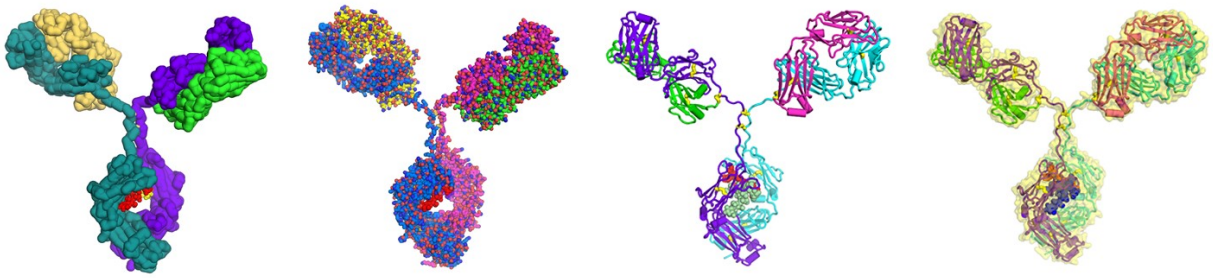


Figure 53. *Simulation of the NIST monoclonal antibody (NIST mAb) Reference Material (SRM 8671).*

Nanoscale Suspension Flow. New instrumentation at NIST's NCNR allows for the study of flow in nanoscale parallelepiped channels using neutron scattering. In support of this study, we are extending the capability of the QDPD dense-suspension code to model similar flows, matching interparticle interactions and flow rates, for comparison to these new scattering experiments. This work will also be extended to include suspensions of monoclonal antibodies and other extended or non-spherical objects in the flow geometry.

[1] N. S. Martys, W. L. George, R. P. Murphy, and K. Weigandt. Certification of SRM 2497: Standard Reference Concrete for Rheological Measurements. NIST Special Publication 260-194, April 2019, 116 pages. DOI: 10.6028/NIST.SP.260-194

[2] ASTM Standard C1874, 2019, New Test Method for Standard Test Method for Measuring the Rheological Properties of Cementitious Materials using a Coaxial Rotational Rheometer. ASTM International, West Conshohocken, PA, 2019.

[3] M. Choi, C. F. Ferraris, N. S. Martys, V. K. Bui, H. R. Trey Hamilton, and D. Lootens. Research Needs to Advance Concrete Pumping Technology. NIST Technical Note 1866, 2015.

[4] S. Z. Jones, D. P. Bentz, N. S. Martys, W. L. George, and A. Thomas. "Rheological Control of 3D Printed Cement Paste." Digital Concrete 2018: First International Conference on Concrete and Digital Fabrication, Zurich Switzerland, September 9-12, 2018.

[5] N. S. Martys, W. L. George, R. P. Murphy, and K. Weigandt. Pipe Flow of Sphere Suspensions having a Power-Law-Dependent Fluid Matrix. *Journal of Rheology* **64** (2020), 445. DOI 10.1122/1.5131021

Monoclonal Antibodies Under High Shear

William George
Nicos Martys (NIST EL)
Steven Satterfield
Judith Terrill

Recent advances at the NIST Center for Neutron Research (NCNR) have enabled neutron scattering experiments for the study of the development of long-range structure and phase transitions under pressure driven flow conditions in a tube geometry. This capability was developed, in part, to study therapeutic proteins, such as monoclonal antibodies, as they flow under high shear rate, such as would be encountered during direct injection via a hypodermic needle.

Such direct injection is desired because it would simplify the delivery of these therapeutic proteins, which are currently only administered intravenously. However, since such injections can lead to shear rates on the order of 10^6/s there is concern that this could lead to structural degradation of the molecules. Even small distortions of the molecular structure could cause aggregation of the protein. The most dangerous potential consequence of such distortion and agglomeration is that the patient may develop an allergic reaction to the medication. Recorded fatalities have occurred where immunogenic response to such aggregates is suspected. For these reasons it is necessary for the pharmaceutical companies to demonstrate that the structure of protein therapeutics is not degrading before high shear rate delivery methods can be utilized.

In support of this research, we have been extending the capabilities of our dense-suspension simulator, QDPD, to match the physical experiments being performed at the NCNR. Recent experiments are in good agreement with preliminary QDPD simulations of similar systems, that is, the flow of suspensions in a tube. The extensions to QDPD will enable the simulation of suspensions of the NIST monoclonal antibody (NIST mAb) Reference Material (SRM 8671) shown in Figure 53. In collaboration with NCNR, the NIST Engineering Lab, and the NIST Material Measurement Lab, we plan

to carry out a suite of simulations designed to determine the effect of the flexibility within monoclonal antibodies on the suspension's viscosity for a set of volume fractions and shear rates like that studied at NCNR. In 2020 we implemented both a spring force and a bending force between the three lobes of the NIST SRM mAb. A third force, controlling the twisting of the lobes will be implemented in the next year.

In addition to the rheological studies, we will also modify QDPD to generate pertinent data from our NIST mAb simulations that will assist in the analysis of neutron scattering data from NCNR experiments on such systems. This is needed since experimental neutron scattering results so far have been shown to be difficult to interpret, primarily due to their complexity.

Early application of this new simulation capability, in collaboration with researchers at the University of Illinois and the University of Delaware, treating the NIST mAb as a rigid protein, will be published in the journal *Structural Dynamics* [1].

[1] Y. Zhai, N. S. Martys, W. L. George, J. Nayem, and Y. Liu. Intermediate Scattering Functions of a Rigid Body Monoclonal Antibody Protein in Solution Studied by Dissipative Particle Dynamic Simulation. *Structural Dynamics*, to appear.

HydratiCA, In Situ Analysis and Machine Learning

Judith Terrill
Justyna Zwolak
Jim Filliben
Jeffrey Bullard (Texas A&M University)

With the extremely long running times and the very large simulations that we now work on, it is not possible to write out the data during the runs. But there are things that need to be studied in the simulations. The only way forward is to do the needed analysis during the runs. The advantage of course, is that we get to do the needed analyses, but the disadvantage is that the analyses need to be done locally in each cell of the simulation and without outside interaction.

We used a large-scale simulation code, HydratiCA [1], to develop and test our techniques. HydratiCA is a stochastic reaction-diffusion model of cement hydration. Hydration transforms paste from a fluid suspension (e.g., cement powder mixed with water) into a hardened solid. This process involves complex chemical and microstructural changes. Understanding and predicting the rates of these changes is a longstanding goal.

HydratiCA integrates a PDE forward in time using small time steps on a 3D lattice of sites, and data analysis is challenging because a single timestep generates a large, many-dimensional data set. Furthermore, there are large parts of that data that are either empty or not interesting to the scientists. The original goal was to try to understand and predict the conditions when a lattice site either runs out of material or space and causes the simulation to greatly reduce the time step, and substantially increase the simulation time.

To address this, we first searched the simulation for a good descriptor (something which can be used to characterize something else: a dimension of representation space). We discovered that reaction saturation index [2, 3]—a quantitative measure of how far a reaction is from equilibrium and also a measure of the thermodynamic driving force for growth—provided valuable information on the state of the simulation at each site. We successfully modeled the saturation index vs. simulation time with Chebyshev polynomials and achieved a dramatic reduction in data because more than a million time steps could easily be reduced to 20 coefficients. This had the additional benefit of smoothing the data which enabled us to calculate reliable first and second derivatives. Now the simulation data could be reduced to a few states. It was then straightforward to find the peaks in the curve that indicated when and where the simulation was running short of resources [4]. The peaks in the saturation index have also been seen in some laboratory experiments [7] and still are not well understood, opening up the possibility of designing laboratory experiments and corresponding simulation runs to do direct comparisons between the experiments and simulation outputs.

Evaluating the Quality of the Fits. Curve fitting goes on in each cell of the simulation. But the goal for real-world runs consists of sizes of the order of 1000^3 cells. Because of the size, it is not possible to look at this data. So, the fits need to be done locally at each cell and without outside intervention. And this means the quality of these fits needs to be evaluated locally. There are recognized ways of doing this. But as Matejka and Fitzmaurice [5, 6] and others have demonstrated, single number statistics are not robust enough for this to be relied upon. The solution is to use images of plots and have a trained classifier "look" at them and report a classification. This year we started assembling a dataset from data accumulated from decades of experiments at NIST.[24] We have created code to analyze each of the datasets and convert them into images. In the coming year, we will extend the dataset. and start building and testing the Deep Learning classifier to provide us with information on the quality of fits.

[24] https://www.itl.nist.gov/div898/software/dataplot/homepage.htm

[1] J. W. Bullard, E. Enjolras, W. L. George, S. G. Satterfield, and J. E. Terrill. A Parallel Reaction-Transport Model Applied to Cement Hydration and Microstructure Development. *Modelling and Simulation in Materials Science and Engineering* **18**:2 (March 2010), 18.

[2] J. W. Bullard, G. W. Scherer, and J. J. Thomas. Time Dependent Driving Forces and the Kinetics of Tricalcium Silicate Hydration. *Cement and Concrete Research* **74** (2015), 26–34.

[3] A. E. Nielsen and J. M. Toft. Electrolyte Growth in Crystals, *Journal of Crystal Growth* **67** (1984), 278-288.

[4] J. E. Terrill. "Curve Fitting with Validation as a First Step to the Discovery of Physical Laws." Workshop III: Validation, Guarantees in Learning Physical Models: from Patterns to Governing Equations to Laws of Nature, Institute for Pure, and Applied Mathematics, UCLA, Los Angeles, CA, October 28, 2019.

[5] A. Cairo, Download the Datasaurus: Never Trust Summary Statistics Alone; Always Visualize Your Data. URL: http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html

[6] J. Matejka and G. Fitzmaurice. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. In *Proceedings of the 2017 ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, CO, May 6, 2017 - May 11, 2017.

[7] J. M. Makar, G. W. Chan, and Y. Esseghaier. A Peak in the Hydration Reaction at the End of the Cement Induction Period. *Journal of Material Science* **42** (2007), 1388–1392.

Towards Robust Autotuning of Noisy Quantum Dot Devices

Joshua Ziegler
Justyna P. Zwolak
Jacob M. Taylor (NIST PML)
Sandesh Kalantre (University of Maryland)
Thomas McJunkin (University of Wisconsin-Madison)
Mark A. Eriksson (University of Wisconsin-Madison)

See full project description on page 101.

Ray-based Classification Framework for Quantum Dot Devices

Justyna P. Zwolak
Jacob M. Taylor (NIST PML)
Sandesh S. Kalantre (University of Maryland)
Thomas McJunkin (University of Wisconsin-Madison)
Brian J. Weber (ShanghaiTech University)

There are myriad quantum computing approaches, each having its own set of challenges to understand and effectively control their operation. Arrays of electrostatically defined quantum dots (QDs) present at the interface of semiconductor devices is one such approach [1]. The current practice of tuning QDs manually or in a semi-automated fashion is a relatively time-consuming procedure, inherently impractical for scaling up device complexity or other integrated applications. Even tuning a double QD constitutes a nontrivial task, with each dot typically being controlled by at least three metallic gates, each of which influences the number of electrons in the dot, the tunnel coupling to the adjacent lead, and the interdot tunnel coupling. At the same time, given the progress in the construction of multi-QD arrays in both one and two dimensions (1D and 2D, respectively) [2,3], it is imperative to replace the current practice of manual tuning to a desirable electronic configuration with a standardized automated method.

As part of our efforts to develop an autonomous tuning system for multi-QD devices, we have recently proposed a new framework for classifying simple high-dimensional geometrical structures: the ray-based classification (RBC) [4]. Rather than relying on the full N-dimensional data tensor, the RBC utilizes a minimal collection of 1D representations, called rays, to assess the geometry of the structure enclosing a given point x_0. The rays are uniquely defined by a set of colinear points

$$\Re(x_0, x_f) = \{x \mid x = (1-t)x_0 + tx_f, \;\; t\epsilon[0,1]\}$$

where x_0 is the ray origin and x_f is the termination point. A sufficiently long ray is expected to intersect the

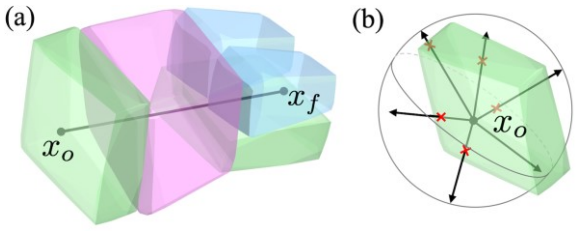


Figure 54. *(a) Visualization of a sample ray $\mathfrak{R}(x_0, x_f)$ from x_0 to x_f in $\mathbb{R}^3$. Different colors represent different classes of polytopes. (b) A sample M-projection centered at x_0. The black arrows represent the M = 6 rays and the red marks indicate the critical features used to determine fingerprints. Adapted from [4].*

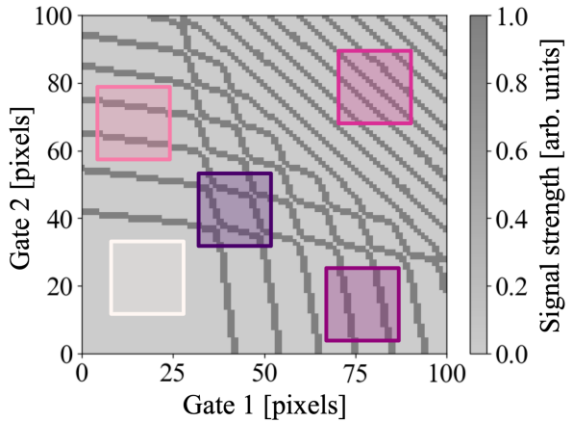


Figure 55. *A sample simulated 2D map showing the different bounded and unbounded polytopes in* $\mathbb{R}^2$.

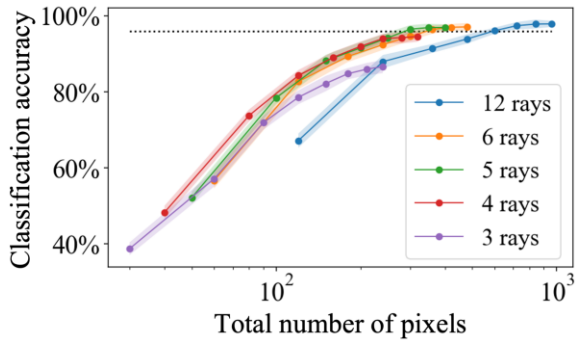


Figure 56. *The RBC performance for varying numbers of rays as a function of the total number of points used to determine the fingerprints. The dashed line represents the benchmark for image-based classification from [6]. Adapted from [4].*

boundary of the structure enclosing x_0 and may cut across more than just one structure (see Figure 54(a) for a depiction). The set of intersection points between the given ray and the structure(s) boundaries, F_{x_0,x_f}, is called the *feature set* in the RBC framework. The natural order in F_{x_0,x_f}, given by the 2-norm distance function allows us to choose a feature nearest to x_0, so-called the *critical feature*.

A collection of M rays of a fixed length r centered at x_0 defines the *M-projection* of x_0. As can be seen in Figure 54(b), it is possible for an M-projection to include rays without any features, that is F_{x_0,x_l}, could in general be empty for $1 \leq \ell \leq M$. Finally, the *point fingerprint* of x_0, which is the primary object of the RBC, is defined as

$$\mathfrak{F}_{x_0} = \left[\Gamma(F_{x_0,x_1}),\ldots,\Gamma(F_{x_0,x_M})\right],$$

where $\Gamma: \{F_{x_0,x_i}\}_1^M \to [0,1]$ is a *weight function* given as

$$\Gamma(F_{x_0,x_i}) = \begin{cases} \gamma\left(x_i^{(c)}\right) & if\ F_{x_0,x_i} \neq 0 \\ 0 & if\ F_{x_0,x_i} = 0 \end{cases}$$

with $x_i^{(c)}$ denoting the critical feature of F_{x_0,x_i} and $\gamma: \mathbb{N}^{>0} \to [0,1]$ being some decreasing function. Importantly, we can prove that if sufficiently many rays in appropriate directions are chosen from x_0, the point fingerprint $\mathfrak{F}_{x_0}$ is sufficient to qualitatively determine the geometry of the convex polytope enclosing x_0 [5]. Formally, the correspondence between the point fingerprint $\mathfrak{F}_{x_0}$ and the set of classes of polytopes filling the $\mathbb{R}^N$ space can be formulated as follows.

> **Ray-based Classification:** Consider a point $x_0 \in \mathbb{R}^N$ and its fingerprint $\mathfrak{F}_{x_0}$. Given a set of bounded and unbounded polytopes filling an N-dimensional space and belonging to C distinct classes, $C \in \mathbb{N}$, determine to which of the classes the polytope enclosing x_0 belongs.

The RBC framework naturally applies to identifying QD states which manifest themselves as distinct geometrical patterns in charge sensor response as a function of the gate voltages, see Figure 55. To test the performance of the fingerprinting approach in differentiating between different states of QD devices, we generated an ensemble of 20 simulated double dot devices that were used to establish a training dataset for a ML algorithm. For each device, we sampled a grid of 1 369 points with 10 and 12 evenly spaced traces per point (a total of 27,380 points). This allowed us to compare performance for five M-projections, with $M = 3, 4, 5, 6,$ and 12. The ray length was varied between 10 and 80 pixels, where the simulation resolution is chosen such that 30 pixels corresponds to the average separation between transition lines. We then trained relatively small deep networks with three fully connected layers (256, 128, and 32 units, respectively) and five classes (i.e., no dot; left, central, and right single dot; double dot). We run 50 training and validation tests per combination of ray's number and length.

The performance of the classifier is shown in Figure 56. As expected, for a fixed number of rays the accuracy of the classifier increases with the total number of points measured. We can also see that increasing the number of rays for a fixed number of points does not necessarily lead to increased accuracy. This is also expected as for a fixed points number and density, increasing the number of rays naturally results in shorter rays. At the same time, rays shorter than the radius of the interior diameter of the shapes result in empty feature sets and uninformative fingerprints. Importantly, we show that the RBC framework preserves the high classification performance while requiring 60 % less data than the traditional image-based classification [6]. A preliminary analysis of the RBC framework using simulated triple-dot system suggest even more significant data reduction.

The RBC is motivated by experiments, particularly those in which sparse data collection is impractical. It

not only reduces the amount of data that needs to be collected, but also can be implemented *in situ* and in an online learning setting, where data is acquired sequentially.

[1] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L.M.K. Vandersypen. Spins in Few-Electron Quantum Dots. *Reviews of Modern Physics* **79**:4 (2007), 1217–1265.

[2] D. M. Zajac, T. M. Hazard, X. Mi, E. Nielsen, and J. R. Petta. Scalable Gate Architecture for a One-Dimensional Array of Semiconductor Spin Qubits. *Physical Review Applied* 6:5 (2016), 054013.

[3] U. Mukhopadhyay, J. P. Dehollain, C. Reichl, W. Wegscheider, and L. M. K. Vandersypen. A 2 × 2 Quantum Dot Array with Controllable Inter-Dot Tunnel Couplings. *Applied Physics Letters* **112**:18 (2018), 183505.

[4] J. P. Zwolak, S. S. Kalantre, T. McJunkin, B. J. Weber, and J. M. Taylor. Ray-Based Classification Framework for High-Dimensional Data. In *Third Workshop on Machine Learning and the Physical Sciences* (ML4PS at NeurIPS 2020), Online, December 11, 2020.

[5] B. J. Weber, S. S. Kalantre, T. McJunkin, J. M. Taylor, and J. P. Zwolak. Theoretical Bounds on Data Requirements for the Ray-Based Classification. In review.

[6] J.P. Zwolak, S.S. Kalantre, X. Wu, S. Ragole, J.M. Taylor. Qflow Lite Dataset: A Machine-Learning Approach to the Charge States in Quantum Dot Experiments. *PLoS ONE* **13**:10 (2018), 1–17.

Ray-Tracing Active Subspace Computations for Quantum Dot Decompositions

Zachary J. Grey
Justyna P. Zwolak
Andrew M. Dienstfrey
Sandesh S. Kalantre (University of Maryland)
Brian J. Weber (ShanghaiTech University)

Electrons confined in arrays of semiconductor nanostructures, called quantum dots (QDs), is one of the many quantum computing approaches. Due to the ease of control of the relevant parameters, fast measurement of the spin and charge states, relatively long decoherence times, and their potential for scalability, QDs are gaining popularity as candidate building blocks for solid-state quantum devices. However, with an increasing number of QD qubits, the relevant parameter space grows exponentially (dimensionality), making heuristic control unfeasible. In semiconductor quantum computing, devices now have tens of individual electrostatic and dynamical gate voltages that must be carefully set to isolate the system to the single electron regime and to realize good qubit performance.

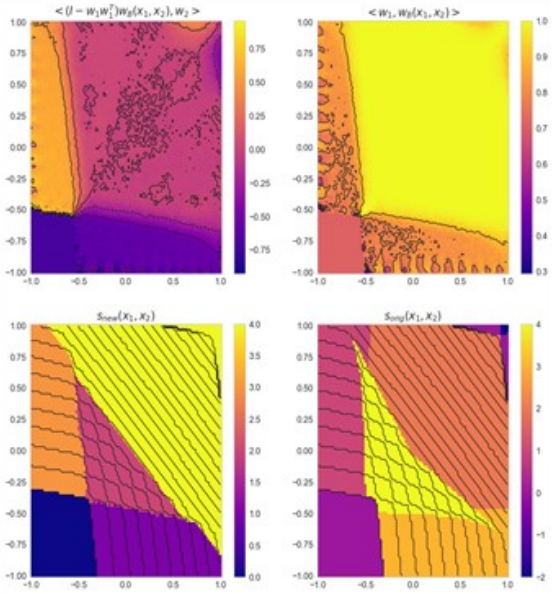


Figure 57. *New domain decomposition (bottom left), local activity criterion (top row), and the original domain decomposition (bottom right) shown with contours of the charge field as black lines (bottom row). Examining changes in the charge field contours (black lines in the bottom plots), the new domain decomposition is a more accurate representation of the device states.*

There has been a recent proposal to use machine learning (ML) techniques as part of the automation process. However, training an ML model requires large amounts of data, preferably with labels indicating the state of the device for a given voltage range. So far, automation efforts rely on datasets that are either labeled manually or come from simulations. However, the former approach does not scale to higher dimensions as the increasing number of QDs lead to too many image features for reliable interpretation. The drawback of the latter approach is a lack of noise typical of real devices. It is thus highly desirable to have automated protocols for decomposing large measurement scans into relevant device states for subsequent interpretation and training of ML models. This work describes an initial explainable approach in two dimensions (two voltage gates) that relies on active subspace decomposition. It will be scaled up to higher-dimensional devices in future work.

In general, given a function $f: \mathcal{X} \subset R^N \rightarrow R$ active subspace analysis proceeds by eigendecomposition of the positive definite, $N \times N$, matrix

$$\boldsymbol{C} = \int_{\mathcal{X}} \nabla f(\boldsymbol{x}) \nabla f(\boldsymbol{x})^T d\mu(\boldsymbol{x}).$$

For $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^{\boldsymbol{T}}$, the *active* subspace is defined by the column span of $\boldsymbol{W}_r = [\boldsymbol{w}_1 \ \ldots \ \boldsymbol{w}_r] \in R^{N \times r}$ where $\boldsymbol{w}_i$ are the first r eigenvectors (active directions) associated

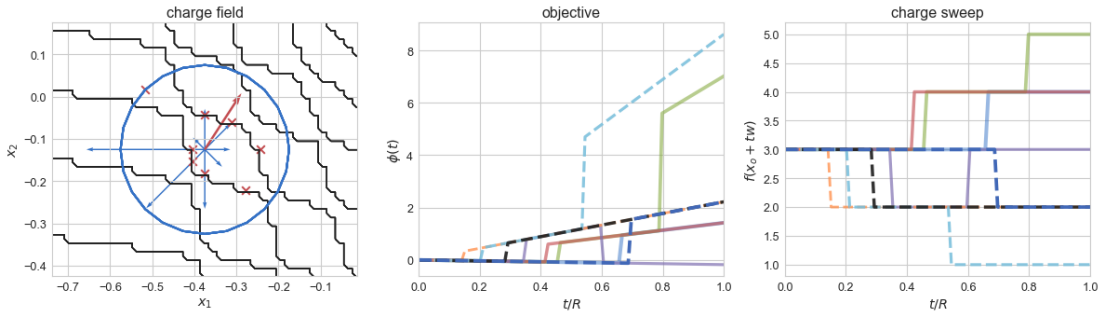


Figure 58. *Visualization of a normalized R-TASC gradient approximation for $N = 2$ dimensional monotonic step function representing a charge surrogate of a quantum dot device. (Left) A visualization of the computed fingerprint. (Center) The objective function responses used to compute features through minimization. (Right) The function response over each of the ray traces. An objective function and 100-point grid search is used to identify changes in the function's response defining the features (red crosses). Seven uniformly distributed rays are traced and each linearly independent ray (four total) is used to build a linear system constituting a kind of second-order approximation of the local activity.*

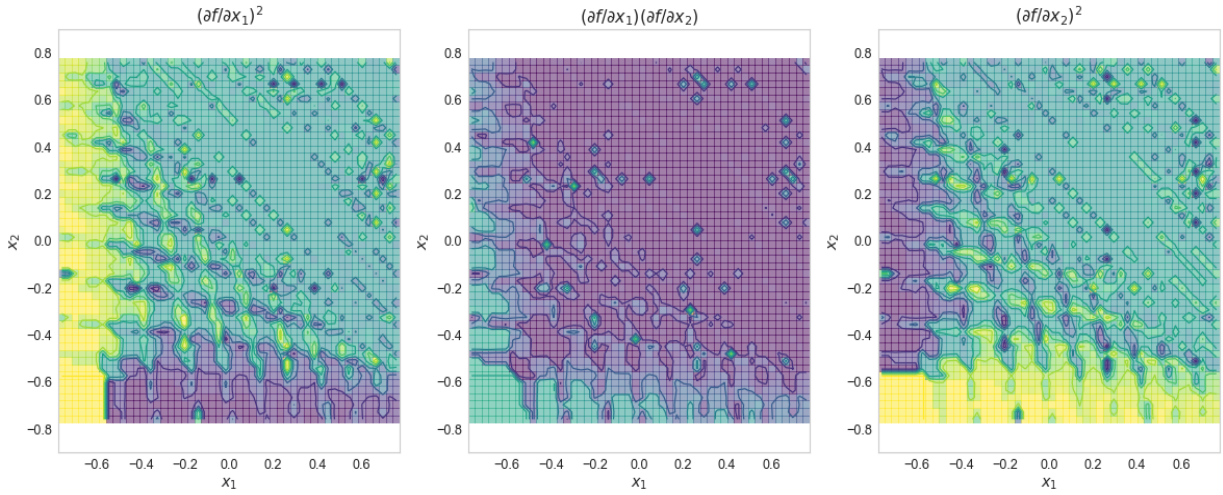


Figure 59. *A visualization of R-TASC product-of-partials fields for $N = 2$ dimensional monotonic step function representing a charge surrogate of a quantum dot device. Local and global integration of this symmetric set of fields will inform the eigenspaces for establishing a set of domain decomposition criteria. Eight uniformly distributed rays are traced with the foot of the ray at each point in a uniform grid sampled over the domain. Again, only the first four linearly independent rays are used to build a linear system constituting a kind of second-order approximation of the local activity.*

with the decaying order of eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. Qualitative observations of two-dimensional QDs (two voltage gates) indicate differences in local and global activity of the monotonic charge field. These differences in activity can be used to establish a set of criteria to inform a domain decomposition. However, special considerations are necessary for the approximation of the integral above over local and global domains given a particular type of charge field measurement.

First, we explored the utility of building criteria from the eigenspaces of a local analog of the $\boldsymbol{C}$ matrix. In this application, f is not differentiable, so we supplement with a surrogate as a local linear model over subdomains, $f(\boldsymbol{x}) \approx \boldsymbol{a}_k^T \boldsymbol{x} + c_k$ for all $\boldsymbol{x} \in \mathcal{X}_k \subset \mathcal{X}$ and $k = 0, \ldots, K$ where $\boldsymbol{a}_k \in R^N$ and $c_k \in R$ for K subdomains. Using this local linear approximation implies

$$\boldsymbol{C}_k \approx \int_{\mathcal{X}_k} \boldsymbol{a}_k \boldsymbol{a}_k^T d\mu(\boldsymbol{x}) \propto \boldsymbol{a}_k \boldsymbol{a}_k^T$$

for a $\boldsymbol{C}_k$ in each subdomain. The linear model constitutes a rank-1 (i.e., $r = 1$) approximation of the matrix over the local subdomain and the eigenspace of such an approximation is nothing more than the span of the unitary direction from the N degree-1 coefficients in the linear model, i.e., the first, and *only*, active direction is $\boldsymbol{w}_k \approx \boldsymbol{a}_k / \|\boldsymbol{a}_k\|_2$ for all $k = 0, \ldots, K$.

Selecting $\mathcal{X}_k$ to be local balls with radius at least as large as the steps in the charge field, we can take approximations of $\boldsymbol{w}_k$ by fitting local linear models to random samples over a discretization of the voltage states. This amounts to an approximation of local activity in the model. That is, we apply a convolution over the field using uniform sampling of a ball of radius 0.2 (a sizing justified by the physics) to fit local linear models restricted to each ball by linear least-squares. The local least-squares fits result in approximations of local directions $\boldsymbol{w}_k$ with the foot of the vector defined at the center of the ball—also known as a local geometric tangent

space. These directions are then contrasted using inner products against a global approximation of the eigenspaces of $\boldsymbol{C}$. Subsequent inner product metrics capture a portion of the salient features in the domain decomposition depicted in the upper row of Figure 57. This response was combined with a half-space criterion to split the remainder of the domain highlighted in yellow depicted in the upper right plot of Figure 57. The resulting domain decomposition after intersecting all criteria is depicted in the bottom left of Figure 57. This procedure was repeated for a variety of simulated quantum dot devices to begin building a database of more accurate domain decompositions (automatic tagging) to be used in subsequent machine learning models.

However, the linear least-squares approach was only viable for use with a simulated device to inform the local linear models. In the lab, measurements cannot be sampled randomly from a local ball to inform local linear least-squares surrogate models. Instead, several ray-based traces are formed and swept over the charge field response. A visualization of a set of rescaled vectors (blue arrows) constituting a candidate set of traces is shown in the left-most plot of Figure 58. The function (charge) response over these traces is depicted in the right-most plot of Figure 58. An objective function for identifying the crossings as rescaled lengths over the rays is depicted in the center plot of Figure 58.

The goal of the ray-based tracing is to compute intersections with the charge field level-sets (denoted as red crosses) in the left-most plot of Figure 58—consistent with the minimums of the objective function responses in the center plot of Figure 58. These computed intersections can be utilized in a forward and backward Taylor series expansion of the function to inform linear system solves which approximate a local analog of activity (a kind of normalized gradient over the ball). In other words, the computed lengths along the rays are proportional to directional derivatives which inform the system

$$P^T \nabla f(\boldsymbol{x}_o) = \boldsymbol{\gamma}(t^+ + t^-)$$

where P represents an $N \times M$, $M \geq N$, matrix aggregating the M linearly independent rays (half of the depicted blue vectors) and $\boldsymbol{\gamma}$ is a signed inverse of the forward and backward lengths, $t^{\pm}$, along each ray. An example solve for $\nabla f(\boldsymbol{x}_o)$ defined at the point $\boldsymbol{x}_o$ (center of the ball) utilizing the ray-based tracing is depicted as the red arrow in the left-most plot of Figure 58. We refer to this method of combining local ray-tracing intersections with level-sets of a function response to infer gradient information as Ray-Tracing Active Subspace Computations (R-TASC).

Finally, repeating the linear solves over a discretized grid $\{(\boldsymbol{x_0})_i\} \subset R^N$ representing a finite domain results in the products-of-partials fields depicted in Figure 59. These products-of-partials constitute the entries of the local and global $\boldsymbol{C}$ matrices which can be combined to inform a similar criterion to the linear least-squares approach. Figure 59 is akin to a "feature space" for establishing a full set of criteria which identify the salient portions of the domain representing different quantum dot states. This is very similar to the linear least-squares approach in that both constitute local linear models. However, any subsequent localized averaging (convolution) over the products-of-partials fields informed by R-TASC are now arbitrary rank as opposed to a rank-1 restriction. Moreover, R-TASC serves as a generalization of an approach for domain decomposition utilizing data measured in the lab. Consequently, future work will focus on methods for real-time domain decompositions of quantum dot devices based on this explainable and interpretable feature space.

Machine Learning Enhanced Dark Soliton Detection in Bose-Einstein Condensates

Justyna P. Zwolak
Shangjie Guo (University of Maryland)
Amilson R. Fritsch (University of Maryland)
Craig Greenberg (NIST ITL)
Ian B. Spielman (NIST PML)

Machine-learning (ML)-based image classification has many applications in science, from particle physics data analysis, dark matter searching, quantum state preparation to material property prediction. In atomic physics, ML has used images of atomic density to locate topological phase transitions and to characterize particles in disordered fields. In this work, we combine convolutional neural networks (ConvNets) with traditional fitting techniques to categorize many-body atomic physics data, and then to extract quantitative information.

Using cold-atom Bose-Einstein condensates (BECs), we focus on solitons, robust solitary waves that retain their size, shape, and speed as they travel. These properties arise from an interplay between nonlinearity and dispersion that is present in many physical systems. Since their first observation in water channels, solitons have been found in rivers and seas, BECs, optical fibers, astronomical plasmas, and even human blood vesicles. Due to their inherent stability, solitons in optical fibers have found commercial applications in long-distance, high-speed transmission lines.

While the natural environment does not allow for the controlled study of quantum solitons, BECs are an excellent medium where individual or multiple solitons can be created on-demand, with all their properties, such as position and velocity, tuned according to necessity [1,

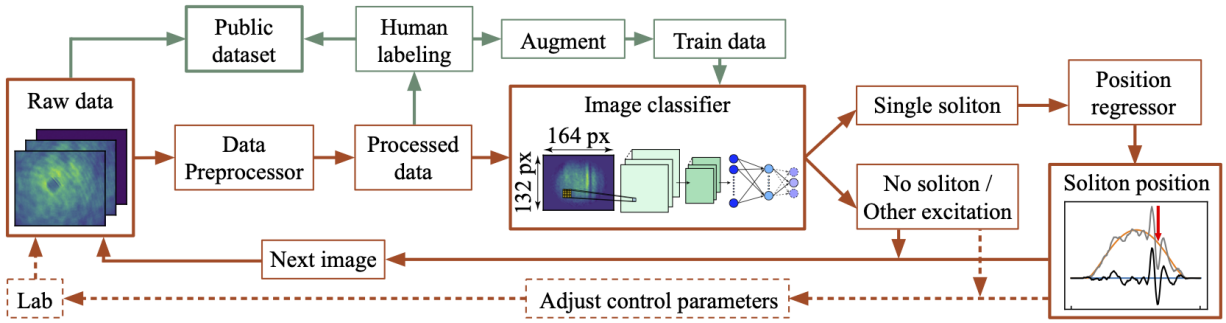


Figure 60. *Schematic of the soliton detection and positioning system. Red boxes and arrows represent the flow of the full system. The dashed red boxes and arrows represent additional components required for a closed-loop implementation. The green boxes and arrows represent additional out-of-loop steps of preparing the classifier and establishing the training dataset.*

2]. Most measurements in BEC experiments produce raw data in the form of images that, in our context, provide information about the solitons' positions within the BEC. The challenge is to efficiently and reliably identify the number of solitons and their locations. Traditional least-squares fitting techniques can locate solitons, provided that the soliton number is known in advance. Currently, the number of solitons is determined manually [1], and this human intervention inhibits the automated analysis of large datasets, which is crucial for solitonic dynamics studies.

Our reliable automated soliton detection and positioning system that takes as input image data and outputs information whether a single soliton is present, and, if so, its location. Since solitons are easily identifiable by human examination of images, this problem naturally connects to the field of computer vision and ConvNet-based image classification. Our algorithm comprises a data preprocessor that converts raw data into a ConvNet-compatible format; a ConvNet image classifier that determines if a single soliton is detected, and a position regressor that locates the soliton within the BEC, when applicable (see Figure 60 for a schematic of the system flow).

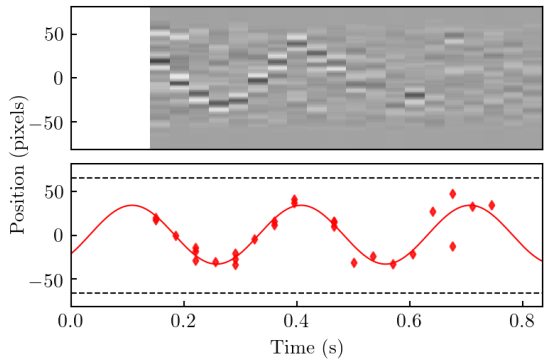


Figure 61. *Oscillation of dark solitons created by applying $1.8(1)\pi$ phase using the standard protocol described in reference [1]. Top panel shows samples for the residuals obtained after subtracting the fit from the 1D profile. Bottom panel shows the soliton positions and sinusoidal fits based on the outputs of the automated system. Dashed lines at 66 pixels represent the edges of the BEC.*

We show that our fully automated system performs comparably to human image classification, autonomously replicating the data analysis in reference [1]. In Figure 61, we show qualitative agreement between experimental data (top row) and our automatically generated results (bottom row). An article reporting this work has recently been published online [3]. In addition, we established a dataset of more than 6000 labeled experimental images of BECs with and without solitonic excitations; this dataset is available via the Midas system at NIST and at data.gov. This dataset provides an opportunity for the data science community to develop more sophisticated analysis tools and to further understand nonlinear many-body physics.

The current performance of our image classifier leaves room for improvement. While tuning the hyperparameters allowed us to improve the initial performance, additional data is necessary to push the limits. However, human labeling is not only time consuming but, as the analysis of the misclassified images revealed, is also not always reliable. Other approaches, such as active learning, may be more suitable for this task. Meanwhile, an enlarged dataset will enable refining the soliton classifier and perform model uncertainty quantification. Together, these refinements may enable reliable in-situ deployment.

Moreover, our study was preconditioned on the assumption of specific structure in the images, leading to our three classes. We will further enlarge our dataset, enabling unsupervised learning strategies to discover additional classes consistent with the data without presumptions. This unsupervised learning of soliton data will be a prototype for ML based discovery with general cold-atom data.

[1] A. R. Fritsch, M. Lu, G. H. Reid, A. M. Piñeiro, and I. B. Spielman. Creating Solitons with Controllable and Near-Zero Velocity in Bose-Einstein Condensates. *Physical Review A* **101** (2020), 053629.

[2] L. M. Aycock, H. M. Hurst, D. K. Efimkin, D. Genkina, H-I Lu, V. M. Galitski, and I. B. Spielman. Soliton Diffusion in Bose–Einstein Condensates. *Proceedings of the National Academy of Sciences* **114** (2017), 2503-2508.

[3] S. Guo, A. R. Fritsch, C. Greenberg, I. B. Spielman, and J. P. Zwolak. Machine-Learning Enhanced Dark Soliton Detection in Bose-Einstein Condensates. arXiv preprint 2101.05404, (2021).

Visualization of Greenhouse Gas Emissions

William Sherman
Judith Terrill
James Whetstone (NIST SPO)
Israel Lopez Coto (NIST SPO)
Anna Karion (NIST SPO)
Kimberly Mueller (NIST SPO)

The NIST Greenhouse Gas (GHG) Measurements Program[25] develops advanced tools and standards for accurately measuring GHG emissions so industries and governments will have the information they need to manage emissions effectively. ACMD's HPCVG is collaborating with James Whetstone, Leader of NIST's Greenhouse Gas Measurements Program, and his team of climate and weather simulation researchers to produce interactive visualizations of their data.

This is HPCVG's first collaboration that solely uses the open-source ParaView visualization toolkit. (see the subsequent project summary for more information on ParaView.) While a necessary part of this project involves data translations between the simulation output and the ParaView tool, the heart of the collaboration is the visualization of climate simulation data in an immersive environment which, by its nature, provides a more intuitive means to see and interact with the data.

We visualized a dataset from a simulation of atmospheric transport of CO_2, CH_4 and CO produced by anthropogenic sources within the area of study (domain of the simulation). Global background concentrations were not included in the simulations. All units are in ppm. The simulation was performed using the Weather Research and Forecasting (WRF) model with passive tracers at 3 km spatial resolution, for a 24-hour period, using Vulcan 3.0, gridded EPA and NEI as emissions inventories for CO_2, CH_4 and CO respectively. See Figure 62 for an example visualization of time step 32 (2:45am EST) in the simulation.

As detailed below, ACMD's HPCVG is working to enhance the immersive capabilities of the ParaView visualization tool. The HPCVG is working on a "plugin" extension to ParaView that enables it to integrate directly into the NIST Immersive Visualization Environment. In addition, ParaView provides an interface to consumer HMD virtual reality systems, which enables another style of immersive interaction with the data. Our ongoing work is to bring these two efforts into convergence making the transition from large-scale systems that are good for group interactions to the more personal HMD systems. In the coming year we will add more capability to our greenhouse gas emissions visualizations, and study them in the CAVE, including the addition of a flexible pre-processing script to better integrate data directly into ParaView.

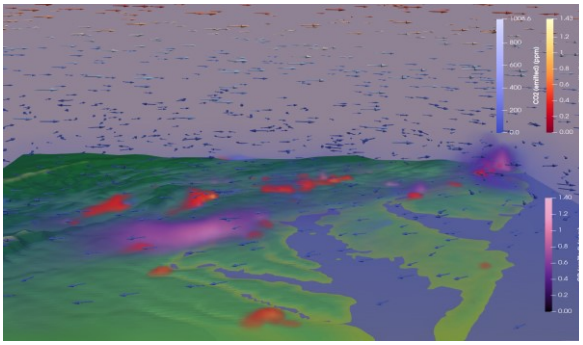


Figure 62. *Visualization of greenhouse gas concentrations of carbon monoxide, carbon dioxide, and methane in the Baltimore Washington corridor. Arrows show the direction and magnitude of the wind with blue slowest and red fastest.*

Transition to Open Source Visualization Software

William George
Terence Griffin
Sandy Ressler
Steven Satterfield
William Sherman
Judith Terrill
Cory Quammen (Kitware)

As part of our research on Virtual Measurements and Analysis, the ACMD High Performance Computing and Visualization Group (HPCVG) operates a fully immersive visualization environment (IVE) and has developed high end visualization (HEV) software to run it. We started developing this software for our IVE more than two decades ago. During this time, we have upgraded and rewritten this software as our understanding scientific visualization in an IVE developed and as outside innovations in hardware appeared. However, there were many limitations to our software. For example, it could only run on one specially configured operating system and it had not kept up with recent advances in hardware visualization capabilities.

Internally this software uses a scene-graph-based framework. Our IVE is on a critical path for the success of collaborations with several NIST research groups and

[25] https://www.nist.gov/greenhouse-gas-measurements

Figure 63. *Tests and measurements are done to ensure the measurements in the virtual world running ParaView match the physical world.*

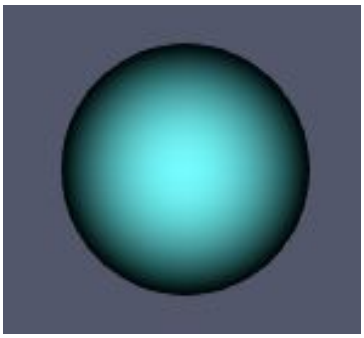


Figure 64. *An example of the new ParaView fatpoint.*

is used at every stage of these collaborations. These projects are diverse and span applications from nanotechnology to medical to materials, and often contribute to standard reference data and materials. For example, the IVE was essential in the success of NIST's development of standard reference materials (SRMs) for the measurement of the flow of cement paste, mortar, and concrete.

To take advantage of recent advances in visualization hardware, we are moving our IVE to ParaView, a fully open source software environment. The ParaView software system is complex. Internally it uses a pipeline and proxy-based framework. The software consists of a Qt interface which uses over 2000 VTK C++ classes to produce visualizations. It runs in an IVE as well as on Windows, Mac OS X, Linux, SGI, IBM Blue Gene, Cray and various Unix workstations, clusters and supercomputers. It supports rendering shaders. It has a new real-time path-tracing back end using NVIDIA RTX technology. ParaView extends the environments that the HPCVG HEV can work in, as well as provides access to real time ray tracing and global illumination made possible with the new GPUs.

This year, we developed configuration files for the NIST IVE that enabled ParaView to work correctly. For this much testing was done. See Figure 63. We created a preliminary NIST SAVG file format (an internal HPCVG format) reader plugin for ParaView that opens SAVG format files and creates data structures in ParaView that enable individual control over line segment and fatpoint visibility. (A fatpoint is a NIST created shader that turns points into spheres with normals for proper display in an IVE. The radius of the spheres is under user control.) The plugin also reads color attribute information that can be used to color the fatpoints.

We also made several important changes to ParaView's virtual reality capabilities. These include changes to core VR infrastructure, user interaction modes, and code documentation and style improvements to aid software maintainability. In terms of core VR infrastructure, we fixed support for the Vrui network protocol support to enable ParaView to read event data from virtual reality devices coming from a Vrui server. We also contributed three new VR interaction modes, two that enable grab actions on points and objects, and one that resets an applied transform. For software maintainability, we thoroughly documented code in ParaView's VR plugin, improved some object naming, and improved several code style elements. We enhanced the VR Plugin with three NIST developed navigation styles. We configured the wand navigation style in ParaView's VR plugin and began evaluation of it in NIST's IVE. A second style that resets the geometry in the world to fit within the CAVE environment has also been created, but not yet configured or evaluated in the NIST IVE. The workflow from ParaView source code changes to deployment on the CAVE has been established. See Figure 64 and Figure 65 for an example ParaView fatpoint and for a ParaView implementation of simulation results of fluid flowing over a mineral sample.

In the coming year we will continue to transfer NIST developed IVE concepts to ParaView VR and continue to move visualization projects to ParaView.

Standards in Visualization

William Sherman
Sandy Ressler
Judith Terrill

Because we have over two decades of experience using immersive virtual reality (IVR) for scientific visualization we have joined the Khronos Group[26] this year in order to be able to contribute to two standards:

[26] https://www.khronos.org

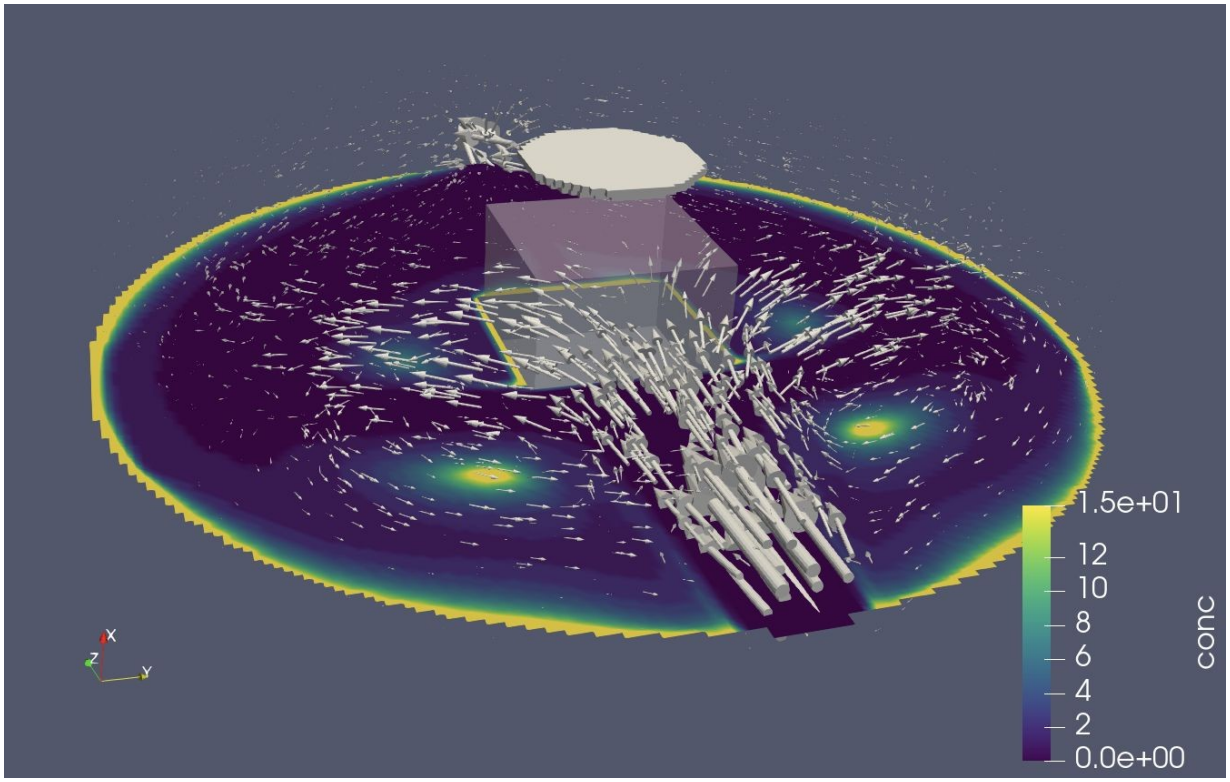


Figure 65. *ParaView implementation of fluid flow simulation results. Arrow glyphs show the direction and magnitude of the flow field. A slice plane pseudo-colored by the concentration data field shows areas of high concentration. The mineral sample is shown as a transparent surface to avoid obscuring the data, and the eyepiece floating at the top of the geometry is displayed as a solid surface for context.*

OpenXR[27] and ANARI[28]. We are now members of both groups. We attend meetings and participate in discussions.

OpenXR. The advent of consumer-ready immersive systems brought about great advances in the capabilities of IVR and other immersive displays while lowering costs considerably. As several companies each pursued their own software ecosystem, many applications were developed that would run exclusively on one vendor's hardware. Subsequently, the Khronos Group has brought together most of the hardware vendors to collaborate on a standard interface between immersive hardware devices and software applications. This standard, known as the OpenXR API, for what is now referred to as "extended reality" (XR), provides a spectrum of interfaces for virtual reality to augmented reality systems.

NIST, and the ACMD High Performance Computing and Visualization Group (HPCVG) in particular, has joined the OpenXR working group to evaluate the developing standard, and to begin exploring the tools available for the beta-stage API. The OpenXR API is designed to be both hardware and operating system agnostic, with implementations for Linux, MacOS, Windows-10, Android and iOS. While many hardware developers will provide a closed-source library through which the API is accessed, there are also open-source versions that can be used, thus preventing vendor lock-in and long-term maintenance issues.

An additional stated goal of OpenXR is to be compatible both with head-worn displays (head-mounted displays), as well as large format, CAVE-style displays such as the NIST Immersive Visualization Environment (IVE). This is one area where the HPCVG team brings expertise to the OpenXR working group.

ANARI. The ANARI (Analytic Rendering Interface for Data Visualization) API is a Khronos Group led effort to provide a common library interface for the rendering of scientific data. Built upon the concepts of the Intel OSPRay renderer, ANARI provides a consistent interface for volumes, point clouds and polygonal data that can be rendered using different methodologies, and tuned to the available graphics hardware of the rendering computer. This can by rasterized on a CPU-based system, or ray-traced renderings on an RTX-level GPU, or anywhere in between.

[27] https://www.khronos.org/openxr

[28] https://www.khronos.org/anari

The ParaView open-source scientific visualization tool, which is increasingly becoming a prevalent analysis tool within the regular HPCVG workflow, presently offers the OSPRay ray-traced rendering library for enhanced rendering. As the ANARI API makes use of the roadmap laid out by OSPRay, many tools presently making use of OSPRay are expected to migrate to the maturing ANARI API. We anticipate ParaView will be one of the many tools that will adopt ANARI, and thus we are participating with the discussions within the Khronos working group.

WebXR Graphics

Sandy Ressler
Matthew Hoehler (NIST EL)
Robert Thompson (NIST SPO)
Justin Slud (University of Maryland)
Matthew Goldman (US Holocaust Memorial Museum)
Jane Klinger (US Holocaust Memorial Museum)
Robert Ehrenreich (US Holocaust Memorial Museum)
Amanda Malinowski (NIST Library)

In the rapidly evolving world of Web based graphics what we used to call WebVR, is now more commonly referred to as WebXR. The change is due to increasing availability not only of Virtual Reality, but Augmented Reality (AR), and Mixed Reality (MR). The term XR connotes all of these types of computer graphics. Our focus remains on prototyping and development of Web based versions of these.

WebVR continues to gain traction in the wider Web. In late 2019 the Immersive Web Working Group of the World Wide Web Consortium (W3C) introduced a new overarching standard that will eventually be supported by all major browser vendors. WebXR encompasses virtual reality (VR), augmented reality (AR), and mixed reality (MR). The W3C has a strong history of introducing standards which go on to be widely and successfully used. Mozilla's Firefox currently supports WebXR and we expect to see widespread adoption. Our work supports this trend by utilizing WebVR and will fully transition to WebXR in the coming year. Web based technologies remain the best method to distribute our work to the public, simplifying the process by eliminating the need for specialized software installations.

Interactive 360 Fire Videos. This continuing area of research, done in collaboration with M. Hoehler of EL, records dramatic fire scenes which the user can see from inside the fire itself. Several cameras were sacrificed to achieve the correct combination of materials and water cooling. The main exemplar fire we are working with is known as the PRICE fire.[29] It was a controlled burn in a

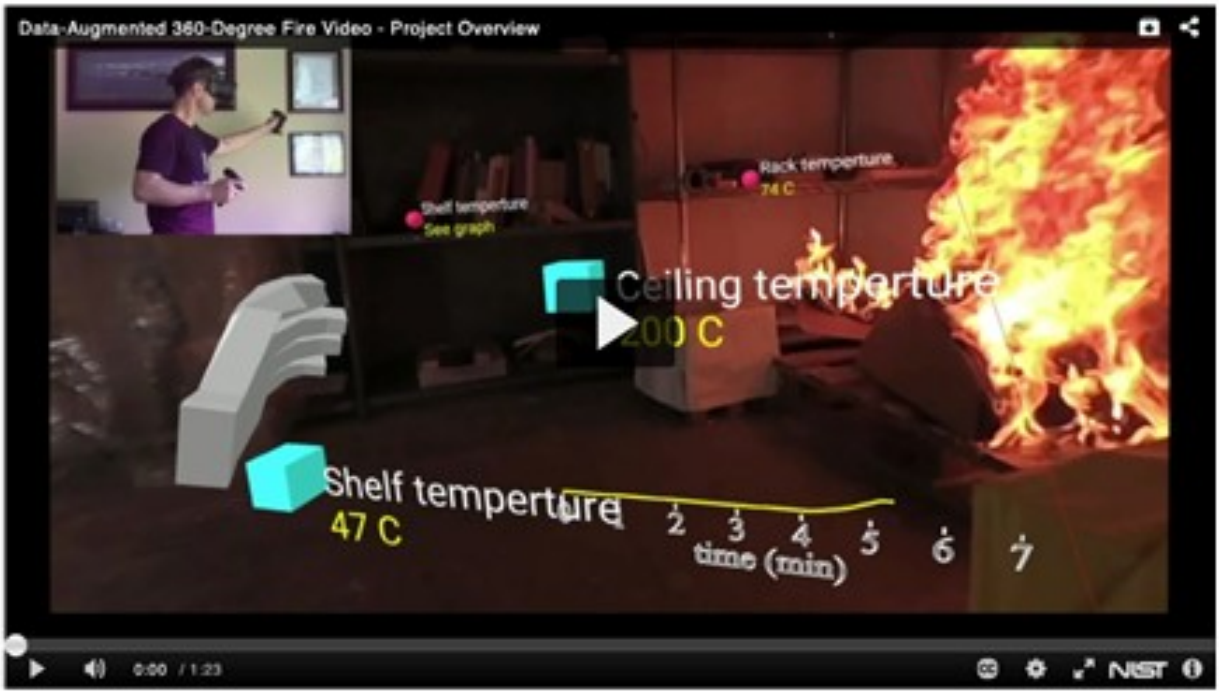


Figure 66. *Manipulating graphs inside WebVR interactive 360-degree fire video.*

[29] https://www.nist.gov/el/fire-research-division-73300/national-fire-research-laboratory-73306/360-degree-video-fire

closet meant to simulate what would happen if a closet at the Smithsonian Institution caught on fire. We successfully extended the work to include the use of VR controllers to virtually grab data graphs and place them anywhere in the scene. The data is synchronized with the 360-degree video presenting users with an immersive and data rich environment. The project was a finalist in the 2020 Auggie awards (best societal impact category), a highly competitive industry event.[30]

DLMF Function Gallery. The DLMF function gallery[31] is in maintenance mode. Given future resources we do intend on extending the WebXR version to include the manipulation of the mathematical functions themselves. We've had discussions with the makers of CalcFlow, an interactive math VR system by Nanome.ai.[32]

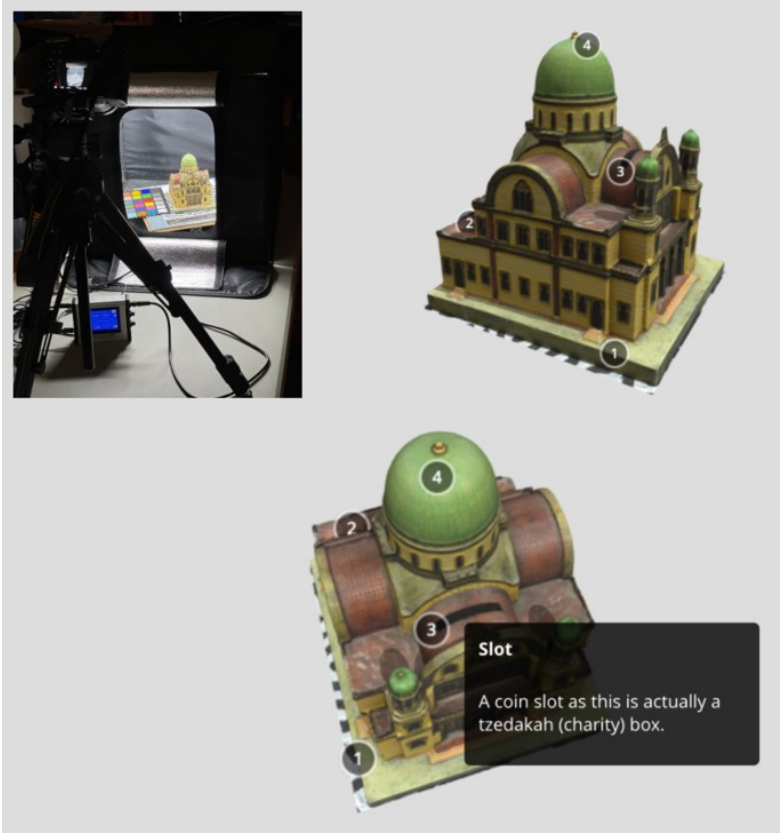


Figure 67. *(Top left) Photogrammetry setup; (Top right) Reconstructed model with viewpoint controls; (Bottom) Annotation content for particular viewpoint.*

3D Object Input and Capture. Viewing a 3D object on a web site of course requires the object itself. Historically we have obtained our objects either from CAD systems or by visualizing data. Missing from that set of objects are representations of "real" objects, artifacts that actually exist in real life. These are items such as museum artifacts, or simply physical objects that are of interest. Allowing curators or the general public to view and manipulate objects that are too precious to handle in real life is an important benefit of access to objects in our virtual world. The goal of this aspect of our WebXR project is to codify standards, processes and techniques to more widely enable dissemination of priceless artifacts and add information about these objects not easily seen via 2D representations. This is not a new domain. Museums and forensic experts have been active in this domain for many years. We are beginning to assemble a working group of experts and end users to make NIST contributions in the context of web accessible objects and interactions with those objects. We've initiated meetings with staff of the United States Holocaust Memorial Museum and the NIST Library (which holds the NIST museum); as well as forensics experts at NIST.

Scanning of objects either via LiDAR (Light Detection and Ranging) and/or photogrammetry are being examined for different applications. The actual scanning of objects is generally a time-consuming error prone activity. New portable LiDAR devices, such as those now available in some high-end smart phones may make 3D scanning much more commonplace. The appropriateness of one technique versus another depends on the application. End users might be museum curators who want the highest accuracy possible or students who want to learn about an artifact. Our work is intended to address and illuminate these practices for as wide a community of users as possible. We remain focused however on the delivery of interactive 3D experiences via the Web.

[30] https://auggies.awexr.com/gallery/xRN-nxNRZ/rJBvXeRJ?search=1bb11ce459d9c740-2

[31] https://math.nist.gov/~SRessler/jas23/MAINPROJECT/gallery.html

[32] https://nanome.ai/calcflow/

Quantum Information

An emerging discipline at the intersection of physics and computer science, quantum information science is likely to revolutionize 21st century science and technology in the same way that lasers, electronics, and computers did in the 20th century. By encoding information into quantum states of matter, one can, in theory, enable phenomenal increases in information storage and processing capability. At the same time, such computers would threaten the public-key infrastructure that secures all of electronic commerce. Although many of the necessary physical manipulations of quantum states have been demonstrated experimentally, scaling these up to enable fully capable quantum computers remains a grand challenge. We engage in (a) theoretical studies to understand the power of quantum computing, (b) collaborative efforts with the multi-laboratory experimental quantum science program at NIST to characterize and benchmark specific physical realizations of quantum information processing, and (c) demonstration and assessment of technologies for quantum communication.

Analog Quantum Algorithms

Lucas Brady
Alexey Gorshkov (NIST PML)
Christopher Baldwin (NIST PML)
Jake Bringewatt (University of Maryland)
Luis-Pedro Garcia-Pentos (University of Maryland)
Yaroslav Kharkov (University of Maryland)
Aniruddha Bapat (University of Maryland)
Ivy Liang (Montgomery Blair High School)

As quantum computers and quantum technologies begin to be realized and become more powerful, it is important to develop algorithms that can exploit these small devices to best utilize any quantum advantage. Such devices are usually noisy and prone to errors, making it difficult to implement many circuit-based quantum algorithms that require a specific sequence of quantum operations. Another class of quantum algorithm that is thought to be more suitable to these small noisy devices is analog quantum algorithms. Rather than being described in terms of discrete quantum operations or gates, these algorithms are described in terms of how the system or Hamiltonian evolve with time.

One of the most promising algorithmic approaches to utilizing small quantum devices involves a variational approach. Here we use classical computation to pick some set of parameters for a quantum calculation or state preparation, using the small quantum computer to sample from the distribution of outputs that results from those parameters. This output is then used to update the classical algorithm, leading to a loop until a desired problem is solved or the state is prepared.

For instance, one of the most prominent quantum variational algorithms is the Quantum Approximate Optimization Algorithm (QAOA) [1]. In this algorithm, a system of quantum bits, or qubits, are exposed to two alternating Hamiltonians. These Hamiltonians are applied for p alternations, with the times for each pulse being fed to the classical computer as its variational parameters. The goal is to approximately prepare the ground state of one of the Hamiltonians (or any other desired state), and the output of sampling the quantum computer for a given parameter set is used in a classical loop to optimize the energy of the produced state (or another desired quantity).

Another important class of analog quantum algorithms is Quantum Adiabatic Optimization (QAO) [2] or its slightly more generalized form, Quantum Annealing (QA). Here, the system is initialized in the ground state of some initial Hamiltonian. If the Hamiltonian changes slowly enough, the quantum adiabatic theorem says that the system will stay in the ground state, allowing us to find the ground state of desired Hamiltonians that might encode some classical problem. These algorithms all use the same ingredients but in different ways. In some work this year [3] (see related feature) we developed a method for finding the optimal protocol using these ingredients. This protocol takes on a bang-anneal-bang form with pulses at the beginning and end and a smooth annealing-like region in the middle. But more than anything, this optimal protocol is just a non-monotonic optimized annealing schedule, and as such it can be used to ask questions beyond the standard annealing framework.

One of those questions is whether there are other Hamiltonians that can be used to prepare the target state other than Hamiltonians that have that target as their ground state. This question is heavily related to questions of controllability of the system and relates to the Lie algebraic structure of the Hamiltonian matrices, and that same Lie algebraic structure relates to Shortcuts to Adiabaticity that allow for faster adiabatic transport (but that are hard to engineer). Using these connections, we [4] designed and justified several alternative Hamiltonians. Comparing them wouldn't work in a standard annealing schedule that would rely on the target state being the final ground state, but using optimized protocols, we could compare everything on equal footing. The result indicated that the original Hamiltonian was indeed

the best, easily constructed, option, but when considering mixtures of multiple Hamiltonians, some of our alternatives proved extremely useful in speeding up the annealing protocols.

The adiabatic theorem provides an upper limit on how long these protocols will take to run, providing a guarantee that a properly designed algorithm can solve the problem in a given time. But that is merely a sufficient condition, not a necessary condition. With many instances, faster computation is possible, with some instances of provable quantum advantage even requiring stricter limits. We began looking into the other side of the scale, examining quantum speed limits. A speed limit is a guarantee that no algorithm can solve the problem in less time than the limit. Combining this with the adiabatic limit, we derived necessary (but not sufficient) conditions for adiabatic transfer, albeit ones that were quite loose. To tighten up results even further and examine the non-adiabatic limit, we have begun examining including specifics about the system of interest, such as locality.

[1] E. Farhi, J. Goldstone and S. Gutmann. A Quantum Approximate Optimization Algorithm. arXiv:1411.4028

[2] E. Farhi, J. Goldstone, S. Gutmann and M. Sipser. Quantum Computation by Adiabatic Evolution. arXiv:0001106

[3] L. T. Brady, C. L. Baldwin, A. Bapat, Y. Kharkov, and A. V. Gorshkov. Optimal Protocols in Quantum Annealing and QAOA Problems. arXiv:2003.08952

[4] I. Liang, J. Bringewatt, A. V. Gorshkov, and L. T. Brady. Quantum Analog Controllability and Alternative Quantum Annealing Hamiltonians. In preparation.

Computational Complexity of Quantum Nonlocality[33]

Carl Miller (NIST ITL)
Honghao Fu (University of Maryland)
William Slofstra (University of Waterloo)

Quantum systems exhibit behaviors that can be very difficult to model with conventional methods of computation. Peter Shor's original quantum factoring algorithm is a striking example of this fact: a hypothetical quantum computer can factor numbers in polynomial time, and yet there is no known way to do the same on a digital computer. Understanding the high levels of computational complexity that occur in quantum physics is a fundamental goal in the science of quantum information.

One of the simplest examples of a quantum phenomenon is the Bell experiment. If a bipartite entangled system is prepared and then measured using local measurements, what possible outcomes can occur? More specifically: what correlations (i.e., conditional probability distributions) can arise from such an experiment? There is, in fact, no digital algorithm that can answer this question (known as the *quantum membership problem* for bipartite correlations) in all cases. In 2017, William Slofstra proved that the problem is undecidable [1].

The goal of the current project is to understand the true sources of computational complexity in the Bell experiment scenario. In [1], Slofstra defined an infinite family of correlations and then proved that there is no algorithm to decide the quantum membership problem in that family. (Specifically, he proved that the halting problem for Turing machines — which is known to be undecidable — can be reduced to the membership problem for correlations in this family.) Crucially, Slofstra used correlations of increasing size, i.e., correlations with an increasing number of measurements in the hypothetical Bell experiment, to achieve this reduction. Recent research on a different quantum problem [2] inspired us to ask a modified question: Is it possible to decide the quantum membership problem for correlations of fixed constant size?

In [3], we proved that the quantum membership problem remains undecidable even when the correlations are of a fixed size. Our proof led us to a surprising conclusion: there is a correlation P such that merely deciding the quantum membership problem among correlations P′ that are arbitrarily close to P (i.e., correlations arising from arbitrarily small adjustments to the probability values of P) is impossible. This result shows how computational complexity can readily occur in seemingly simple scenarios. Our result is connected to a long line of groundbreaking results from the last several years on the computational power of quantum nonlocality, including the recent capstone result MIP*=RE [4].

Notably, papers on this topic — including ours — have relied substantially on tools from quantum self-testing. It is known that there are certain correlations that can only occur if the Bell equipment uses a unique entangled state and a unique set of measurements. This fact (quantum *self-testing* or *self-checking*) was observed by Mayers and Yao many years ago [5] and was positioned as a technique for proving security for quantum cryptography. In our work it fulfills a different purpose: it allows us to encode a Turing machine into the coordinates of a constant-sized correlation matrix, thus enabling a new version of Slofstra's undecidability proof. This serves as an example of the (sometimes surprising) synergy between different problems in quantum information science.

[33] This work was performed by the ITL Computer Security Division. It is reported here to provide a complete picture of the ITL Quantum Information Science program.

Figure 68. *QuICS Fellow Carl Miller of the ITL Computer Security Division (right) discusses quantum protocols with QuICS postdocs and students. Photo courtesy of the University of Maryland.*

[1] W. Slofstra. The Set of Quantum Correlations Is Not Closed. *Forum of Mathematics, Pi* **7** (2019), e1.

[2] J. Bausch, T. S. Cubitt, A. Lucia, and D. Perez-Garcia. Undecidability of the Spectral Gap in One Dimension. *Physical Review X* **10** (2020), 031038.

[3] H. Fu, C. A. Miller, and W. Slofstra. The Membership Problem for Constant-sized Quantum Correlations Is Undecidable. Preprint, arXiv:2101.11087 (2021).

[4] Z. Ji, A. Natarajan, T. Vidick, J. Wright, and H. Yuen. MIP* = RE. Preprint, arXiv:2001.04383 (2020).

[5] D. Mayers, and A. Yao. Quantum Cryptography with Imperfect Apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS)*, 1998, 503-509.

Computations with Greater Quantum Depth Are Strictly More Powerful (Relative to an Oracle)

Matthew Coudron
Sanketh Menda (University of Waterloo, Canada)

A popular paradigm in algorithm design for near-term quantum computers is to extend the limited capabilities of current quantum hardware by interleaving quantum and classical computation in a "hybrid quantum algorithm." This paradigm includes the variational quantum algorithm often used in quantum chemistry, and the popular quantum approximate optimization algorithm (QAOA). But, how much computational power should we expect these approaches to have, as compared to a full, idealized, quantum computer? It turns out that the answer to this fundamentally important question is not entirely understood and has been the subject of conjectures and open problems in quantum computing for more than 15 years. In this research project we specifically consider the case when the near-term quantum computers in question are limited in terms of their circuit depth. Our main technical accomplishment is to resolve a conjecture of Aaronson by exhibiting the first proof that there is a concrete computational problem which can be solved efficiently by a full scale quantum computer, but not by a hybrid quantum-classical algorithm in which the quantum component has low circuit depth.

It is worth stating here the two most fundamental conjectures on this topic, which, in a sense, point in opposite directions (although they are technically incomparable). A conjecture of Jozsa [4] states that any polynomial-time quantum computation can be simulated by polylogarithmic-depth quantum computation interleaved with polynomial-depth classical computation. Separately, Aaronson conjectured [1] that there exists an oracle O such that $BQP^O \neq BPP^{BQNC^O}$. That is, that there exists an oracularized computational problem that can be solved by full scale quantum computation (BQP), but not by classical and low-depth quantum computation interleaved with each other (BPP^{BQNC}). These conjectures are intriguing allusions to the unresolved potential of combining classical and low-depth quantum computation. Note that Jozsa's conjecture is plausible because even some of the most sophisticated quantum algorithms, such as the quantum factoring algorithm, are known to be implementable with logarithmic depth quantum circuits and classical post-processing, as shown in [3].

In this work we show that the welded tree problem, which is an oracle problem that can be solved in quantum polynomial time as shown by Childs et al. [2], cannot be solved in BPP^{BQNC}, nor can it be solved in the class that Jozsa describes. This proves Aaronson's conjecture and provides a counterpoint to Jozsa's conjecture relative to the welded tree oracle problem. More precisely, we define two complexity classes, HQC and JC whose languages are decided by two different families of interleaved quantum-classical circuits. HQC contains BPP^{BQNC} and is therefore relevant to

Aaronson's conjecture, while JC captures the model of computation that Jozsa considers. We show that the welded tree problem gives an oracle separation between either of {JC, HQC} and BQP. Therefore, even when interleaved with arbitrary polynomial-time classical computation, greater "quantum depth" leads to strictly greater computational ability in this relativized setting.

[1] S. Aaronson. Ten Semi-Grand Challenges for Quantum Computing Theory. Blog post, July 2005. URL: https://www.scottaaronson.com/writings/qchallenge.html

[2] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman. Exponential Algorithmic Speedup by a Quantum Walk. *In Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, San Diego, CA, June 9-11, 2003, 59–68. DOI: 10.1145/780542.780552

[3] R. Cleve and J. Watrous. Fast Parallel Circuits for the Quantum Fourier Transform. In *41st Annual Symposium on Foundations of Computer Science* (FOCS 2000), Redondo Beach, CA, November 12-14, 2000, 526–536. DOI: 10.1109/SFCS.2000.892140

[4] R. Jozsa. An Introduction to Measurement Based Quantum Computation. NATO Science Series, III: Computer and Systems Sciences. *Quantum Information Processing-From Theory to Experiment* **199** (2006), 137–158. URL: https://arxiv.org/abs/quant-ph/0508124

Quasi-polynomial Time Approximation of Output Probabilities of Constant-depth, Geometrically Local Quantum Circuits

Matthew Coudron
Nolan J. Coble (University of Maryland)

It is believed that quantum computers will provide an exponential computational speed-up over classical machines for certain computational tasks. Exactly which computational tasks have this potential, and how far the near-term quantum hardware can go towards this goal, are two of the biggest open questions about quantum computation today. In this research project we study the classical complexity of simulating certain near-term quantum computational models. We do not assume a "brute-force" simulation technique, but rather, make use of non-trivial structure within the near-term quantum circuits to achieve provable simulation results in sub-exponential time. These results can serve as a litmus test for determining which near-term quantum circuit architectures have the potential for an exponential quantum advantage. If an architecture can be classically simulated efficiently, then it does not have the potential for an exponential quantum advantage. On the other hand, these novel simulation results may provide some intuition for the conditions under which an exponential advantage might be achieved by quantum circuits.

As our central technical contribution, we present a classical algorithm that, for any 3D geometrically-local, constant-depth quantum circuit C, and any bit string $x \in \{0,1\}^n$, can compute the quantity $\left|< x|C \mid 0^{\otimes n} >\right|^2$ to within any inverse-polynomial additive error in quasi-polynomial time. It is known that it is #P-hard to compute this same quantity to within 2^{-n^2} additive error [3]. The previous best-known algorithm for this problem used $O(2^{n^{\frac{1}{3}}}\left(\frac{1}{\epsilon}\right)^2)$ time to compute probabilities to within additive error ϵ [1]. Notably, the paper [1] included an elegant polynomial time algorithm for the same estimation task with 2D circuits, which makes a novel use of 1D matrix product states (MPS) carefully tailored to the 2D geometry of the circuit in question. Surprisingly, it is not clear that it is possible to extend this use of MPS to address the case of 3D circuits in polynomial time. This raises a natural question as to whether the computational complexity of the 3D problem might be drastically higher than that of the 2D problem. In this work we address this question by exhibiting a quasi-polynomial time algorithm for the 3D case. We believe that our algorithm extends naturally to any fixed dimension D by induction on the dimension, but we focus on the 3D case, as the simplest unresolved case, for concreteness. Furthermore, we show that, under a natural, polynomial-time-checkable condition on the circuit C, our algorithm runs in polynomial time. This highlights the possibility that the super-polynomial worst-case time bound on our algorithm might be due to limitations of our analysis. In order to surpass the technical barriers encountered by previously known techniques we are forced to pursue a novel approach: Constructing a recursive sub-division of the given 3D circuit using carefully designed block-encodings.

Our algorithm has a divide-and-conquer structure, demonstrating how to approximate the desired quantity via several instantiations of the same problem type, each involving 3D-local circuits on at most half the number of qubits as the original. This division step is then applied recursively, expressing the original quantity as a weighted sum of smaller and smaller 3D-local quantum circuits. A central technical challenge is to control correlations arising from the entanglement that may exist between the different circuit "pieces" produced this way. We believe that the division step, which makes a novel use of block-encodings [2], together with an inclusion-exclusion style argument to reduce error, may be of interest for future research on low depth quantum circuits.

[1] S. Bravyi, D. Gosset, and R. Movassagh. Classical Algorithms for Quantum Mean Values. QIP, 2020. URL: https://arxiv.org/abs/1909.11485

[2] A. Gilyen, Y. Su, G. Hao Low, and N. Wiebe. Quantum Singular Value Transformation and Beyond: Exponential Improvements for Quantum Matrix Arithmetics. STOC, 2019. URL: https://arxiv.org/pdf/1806.01838.pdf

[3] R. Movassagh. Quantum Supremacy and Random Circuits. QIP, 2020. URL: https://arxiv.org/pdf/1909.06210.pdf

Trading Locality for Time: Certifiable Randomness from Low-Depth Circuits

Matthew Coudron
Jalex Stark (University of California at Berkeley)
Thomas Vidick (Caltech)

The generation of certifiable randomness is one of the most fundamental information-theoretic tasks that meaningfully separate quantum devices from their classical counterparts. We propose a protocol for exponential certified randomness expansion using a single quantum device. The protocol calls for the device to implement a simple quantum circuit of constant depth on a 2D lattice of qubits. The output of the circuit can be verified classically in linear time and is guaranteed to contain a polynomial number of certified random bits assuming that the device used to generate the output operated using a (classical or quantum) circuit of sub-logarithmic depth. This assumption contrasts with the locality assumption used for randomness certification based on Bell inequality violation and more recent proposals for randomness certification based on computational assumptions. Furthermore, to demonstrate randomness generation it is sufficient for a device to sample from the ideal output distribution within constant statistical distance.

Our procedure is inspired by recent work of Bravyi et al. [1], who introduced a relational problem that can be solved by a constant-depth quantum circuit, but provably cannot be solved by any classical circuit of sub-logarithmic depth. We develop the discovery of Bravyi et al. into a framework for robust randomness expansion [2]. Our results lead to a new proposal for a demonstrated quantum advantage that has some advantages compared to existing proposals. First, it does not rest on any complexity-theoretic conjectures but relies on the physical assumption that the adversarial device being tested implements a circuit of sub-logarithmic depth. Second, success on our task can be easily verified in classical linear time. Finally, our task is more noise-tolerant than most other existing proposals that can only tolerate multiplicative error, or require additional conjectures from complexity theory; in contrast, we are able to allow a small constant additive error in total variation distance between the sampled and ideal distributions.

[1] S. Bravyi, D. Gosset, and R. König. Quantum Advantage with Shallow Circuits. *Science* **362**:6412 (2018), 308-311.

[2] M. Coudron, J. Stark, and T. Vidick. Trading Locality for Time: Certifiable Randomness from Low-Depth Circuits. *Communications in Mathematical Physics*, to appear.

Post-Quantum Cryptography

Yi-Kai Liu
Gorjan Alagic (NIST ITL)
Jacob Alperin-Sheriff (NIST ITL)
Daniel Apon (NIST ITL)
Lily Chen (NIST ITL))
David Cooper (NIST ITL)
Quynh Dang (NIST ITL)
John Kelsey (NIST ITL)
Carl Miller (NIST ITL)
Dustin Moody (NIST ITL)
Rene Peralta (NIST ITL)
Ray Perlner (NIST ITL)
Angela Robinson (NIST ITL)
Daniel Smith-Tone (NIST ITL)

Since 2016, NIST has been engaged in a formal process to develop standards for post-quantum cryptography (PQC). The goal is to standardize one or more cryptosystems that could replace those currently used schemes that are vulnerable to attack by quantum computers. These include RSA, Diffie-Hellman and elliptic curve cryptosystems, which play a crucial role in electronic commerce and cybersecurity. While large quantum computers have not yet been built, NIST believes it is prudent to begin preparing for that possibility.

NIST is focusing on three main functionalities: public-key encryption, key establishment, and digital signatures. These are fundamental cryptographic primitives that enable a variety of applications, including secure web browsing, digital certificates, and secure software updates.

There are a number of candidate cryptosystems that are believed to be quantum secure. These are based on a variety of mathematical techniques, including high-dimensional lattices, coding theory, systems of multivariate polynomial equations, elliptic curve isogenies, hash-based signatures, and many others. However, further research is needed in order to increase confidence in the security of these schemes, and to improve their practical performance.

In July 2020, NIST began the third round of the analysis and evaluation of candidate cryptosystems for standardization [1]. From the 26 candidates in the second round of the evaluation process, NIST selected a smaller number of candidates to move onto the third round. These consisted of seven “finalists” (four

schemes for public-key encryption and key establishment, and three schemes for digital signatures) that will be considered for standardization at the end of the third round; and eight "alternates" (five schemes for public-key encryption and key establishment, and three schemes for digital signatures) that will still be considered for possible standardization at a later date.

The selection of these third-round candidates was informed by a large amount of research and analysis by the cryptography community, as well as internal discussions at NIST. During the second round of the evaluation process, there was substantial progress on several important issues, including checking the validity of security proofs, verifying the estimates of the concrete security of the different schemes, developing new cryptanalytic attacks, finding potential vulnerabilities (such as side-channel attacks), developing countermeasures to known vulnerabilities, developing and benchmarking high-quality software implementations of these cryptosystems on a variety of computing platforms, and assessing the practical feasibility of deploying these cryptosystems in various use-cases.

By selecting a small number of finalist candidates, NIST has also chosen some likely paths to the development of a PQC standard. Many of these paths rely heavily on lattice-based cryptography, using lattices that have algebraic structure. These techniques have good performance and are believed to be secure. However, NIST recognizes that research on PQC is ongoing, and there is a risk that a future advance in cryptanalysis could change the situation significantly. By choosing alternate candidates, NIST is attempting to hedge against this risk.

NIST expects the third round of the evaluation process to continue for 12-18 months. During this time, NIST hopes to solidify its understanding of the remaining candidates. In June 2021, NIST plans to host a Third NIST PQC Standardization Conference, and by early 2022, NIST expects to select a small number of candidates for standardization.

[1] G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, J. Kelsey, Y.-K. Liu, C. Miller, D. Moody, R. Peralta, R. Perlner, A. Robinson, and D. Smith-Tone. Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process. NISTIR 8309, July 2020. DOI: 10.6028/NIST.IR.8309

Quantum Information Science

Scott Glancy
Emanuel Knill
Sae Woo Nam (NIST PML)
Konrad Lehnert (NIST PML)
Tasshi Dennis (NIST PML)
Ezad Shojaee
Arik Agavayan
Mohammad Alhejji
Shawn Geller
Alex Kwiatkowski
Karl Mayer
James R. van Meter
Curtis Rau (University of Colorado)
Akira Kyle (University of Colorado)
Lynden K. Shalm (University of Colorado)
Yanbao Zhang (NTT Corporation)
Gerardo Ortiz (Indiana University)
Hong Hao Fu (University of Maryland)

Quantum computing and communication technology is advancing rapidly. There are now several commercial services providing access to quantum registers with dozens of physical qubits. Quantum communication that can connect quantum computers is not far behind, with demonstrations of stationary to flying qubit conversion for various platforms coming at a rapid pace. ACMD researchers are involved in a broad range of research contributing to these developments. Projects include theoretical and experimental efforts to demonstrate practical and device-independent randomness generation and expansion, mathematical analyses of fault-tolerant control for exchange-based quantum computing and analysis of tolerable optical noise for quantum transduction in relevant quantum network configurations.

Randomness Generation. NIST capabilities for running loop-hole-free Bell test experiments provide an opportunity to demonstrate device-independent randomness generators whose randomness is certified even if the quantum devices used have been manufactured by untrusted parties. ACMD researchers have developed the most data-efficient protocols for such randomness generators and have collaborated with PML researchers to demonstrate them. This collaboration resulted in two successful demonstrations. One is the first randomness generator that can rapidly and repeatedly produce blocks of 512 bits of randomness with strong certificates sufficient for use in secure communication protocols [1]. The other demonstration is of randomness expansion, whereby a source of seed randomness is expanded into a higher-rate source of device-independent randomness [2]. The theory underlying these demonstrations has appeared in [3]. There is ongoing work to improve the randomness generation devices and to reduce the resources required for randomness expansion.

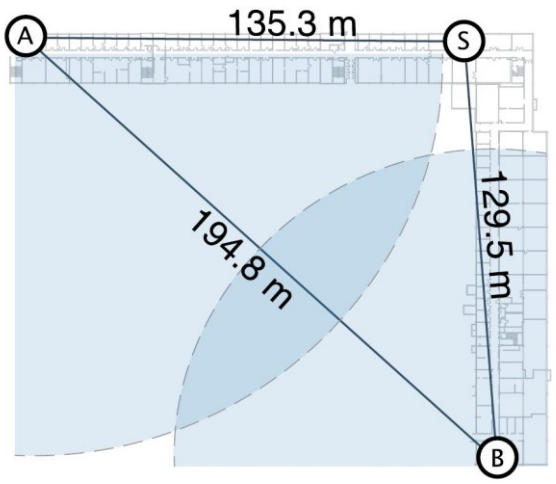


Figure 69. *Schematic of the Bell test layout superimposed on the building. Locations of the Bell test stations A and B and the entangled photon source S superimposed on a building schematic. The shaded circles represent how far local information about A's (B's) measurements could have propagated at the speed of light when B (A) have completed their measurement for a trial. A's and B's settings choices and measurement outcomes are therefore space-like separated.*

Fault-Tolerant Quantum Computing. Low-level protection against errors during operation of physical qubits is required to support fault-tolerant operation of quantum computers. One source of errors for co-located qubits is collective noise. This noise can be protected against by encoding logical qubits in decoherence-free subsystems. Quantum gates can then be implemented with exchange interactions. A system that naturally supports these operations consists of electrically defined quantum dots being investigated as candidate qubits for quantum computation. ACMD researchers have been investigating the mathematical properties of the relevant decoherence-free subsystems supported by quantum systems of all dimensions. They have essentially solved the problem of classifying collections of decoherence-free subsystems for which the exchange interaction is universal without ancillas and demonstrated that a moderately sized ancilla system suffices for all collections. Next steps include determining practical schemas to implement specific universal gate sets.

Quantum Networks. ACMD researchers are participating in multiple quantum network projects involving conversion between stationary qubits (superconducting or ion) and optical qubits. This conversion is a necessity for linking quantum computers into quantum networks.

There are multiple protocols for using such conversion to establish entanglement between remote computers. ACMD researchers analyzed Gaussian entanglement swapping schemes, which are among the simplest to implement. They determined the combinations of efficiency and added noise parameters for which such a scheme can be successful. Because the average efficiency of the sources needs to be at least 1/2 and because of the destructiveness of the added noise, the planned entanglement generation configuration in one of the projects needed to be changed. Work is continuing on characterizing the practical entanglement generation configurations and their tolerance to efficiency, added noise and other unwanted effects.

[1] Y. Zhang, L. K. Shalm, J. C. Bienfang, M. J. Stevens, M. D. Mazurek, S. W. Nam, C. Abellán, W. Amaya, M. W. Mitchell, H. Fu, C. A. Miller, A. Mink, and E. Knill. Experimental Low-Latency Device-Independent Quantum Randomness. *Physical Review Letters* **124** (2020), 010505. DOI: 10.1103/PhysRevLett.124.010505

[2] L. K. Shalm, Y. Zhang, J. C. Bienfang, C. Schlager, M. J. Stevens, M. D. Mazurek, C. Abellán, W. Amaya, M. W. Mitchell, M. A. Alhejji, H. Fu, J. Ornstein, R. P. Mirin, S. W. Nam, and E. Knill, Device-Independent Randomness Expansion With Entangled Photons. *Nature Physics*, to appear. Preprint: https://arxiv.org/abs/1912.11158

[3] Y. Zhang, H. Fu, and E. Knill, Efficient Randomness Certification by Quantum Probability Estimation. *Physical Review Research* **2** (2020), 013016. DOI: 10.1103/PhysRevResearch.2.013016

Quantum Characterization Theory and Applications

Scott Glancy
Emanuel Knill
David T. C. Allcock (NIST PML)
José Aumentado (NIST PML)
Katarina Cicak (NIST PML)
Thomas Gerrits
Dietrich Leibfried (NIST PML)
Adriana E. Lita (NIST PML)
Richard P. Mirin (NIST PML)
Sae Woo Nam (NIST PML)
Raymond W. Simmonds (NIST PML)
Daniel Slichter (NIST PML)
John D. Teufel (NIST PML)
Andrew Wilson (NIST PML)
Arik Avagyan
Italo Bezerra (Federal University of Ceará)
Peter Bierhorst (University of New Orleans)
Shaun C. Burd (University of Colorado)
Mingyuan Hong (Louisiana State University)
Shawn Geller
Hannah Knaack (University of Colorado)
Steve Kolthammer (Imperial College London)
Shlomi Kotler (University of Colorado)
Alex Kwiatkowski
Florent Lecocq (University of Colorado)
Omar S. Magaña-Loaiza (Louisiana State University)
Gabriel A. Peterson (University of Colorado)
David Phillips (University of Oxford)

Ezad Shojaee
R. T. Sutherland (Lawrence Livermore National Lab)
Raghavendra Srinivas (University of Colorado)
Hilma Vasconcelos (Federal University of Ceará)
David J. Wineland (University of Oregon)
Chenglong You (Louisiana State University)

Many emerging technologies will exploit quantum mechanical effects to enhance metrology, computation, and communication. Developing these technologies requires improved methods to characterize the performance of quantum devices. This characterization requires solving statistical problems such as estimating an underlying quantum state, measurement, or process by using a collection of measurements made on the quantum system. Alternatively, one may also want to estimate figures-of-merit such as fidelity, error rates, and entanglement measures from that data. Accurate quantum characterization allows experimentalists to answer questions like "What is happening in my quantum experiment?" or "How well will my system perform some quantum information protocol?" and to characterize uncertainty in that answer.

NIST's Ion Storage Group has pioneered one of the world's most successful quantum computer development projects. To keep pace with their recent advances in qubit preparation, logical operation, and measurement fidelities, ACMD researchers are developing more advanced statistical techniques to characterize trapped-ion quantum computers. An ion qubit is measured by counting the number of photons the ion releases when illuminated by a fluorescent laser. A random number of photons will be produced, but the probability distribution depends on the ion's state. Distinguishing the $|0\rangle$ and $|1\rangle$ states requires differentiating between the two, slightly overlapping, probability distributions. Also, the ion might change its state during the measurement. ACMD researchers have developed a hidden Markov model of the fluorescent measurement process, a strategy for manipulating the ion during the measurement based on the incoming photon detections, and a Bayesian system for inferring the hidden Markov model's parameters and the ion's state at the start of the measurement process. These tools enable higher fidelity qubit measurement and a better understanding of measurement error [1].

One difficulty with scaling-up trapped-ion quantum computers is the management of lasers used to apply logic gates. To overcome this difficulty, the Ion Storage Group has developed techniques to apply all logic gates using only microwave fields (see Figure 70). These microwave logic gates were used in an experiment demonstrating universal quantum logic operations on two qubits. ACMD researchers designed a strategy for estimating the fidelity with which the microwave gates could entangle the qubits. That fidelity is in the interval

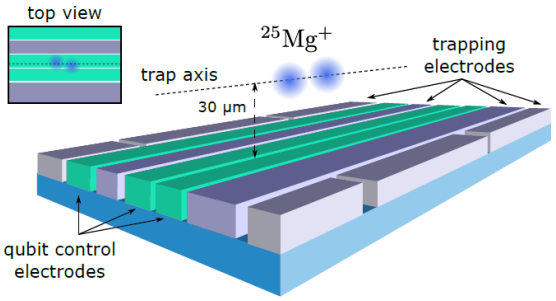


Figure 70. *3D perspective view of central region of surface electrode ion trap (not to scale). For this experiment, two 25Mg+ ions are trapped 30 μm above the trap surface. MHz and GHz currents used to generate the entangling interaction are driven along the qubit control electrodes (green), while oscillating (purple) and static voltages (gray) are applied to the trapping electrodes to create the confining potential.*

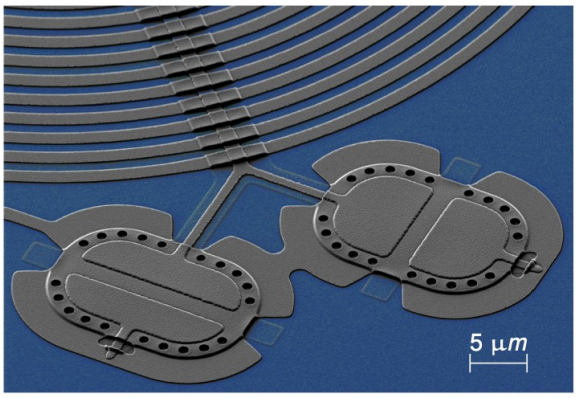


Figure 71. *Scanning electron micrograph of drum. Two aluminum drums are suspended above a sapphire substrate (false colored in blue) resulting in well-defined harmonic modes in the direction perpendicular to the substrate. Each drum forms the top plate of a capacitor, along with the bottom plate which is fixed to the substrate. The oscillations of the drums are entangled through their interaction with a microwave-frequency LC circuit.*

[0.9983, 1] with 68 % confidence [2], which rivals the best fidelities achieved with lasers.

Inspired by the needs of the Metrology with Interacting Photons Innovations in Measurement Science program, ACMD researchers have developed a generalization of traditional homodyne detection. Traditional homodyne detection requires a strong reference beam and can only estimate the state in the mode matching the reference beam, but our generalization can use a weak reference beam (as required for some integrated circuit designs) and can learn about correlations between the signal mode and photons in nearby modes [3]. ACMD researchers have partnered with PML to demonstrate this measurement strategy in an optical experiment. Preliminary analysis shows high-fidelity results.

A basic resource state for continuous-variable quantum information processing is a low-temperature squeezed state. Such states are often prepared in optical

modes or in the motion of trapped ions. ACMD researchers have developed a method to estimate the amount of squeezing and temperature from measurements of the number of photons or phonons in the state [4]. They are using this method to study the squeezing of ion motion, which was previously characterized assuming that its temperature was 0.

One important test of quantum theory is verification of entanglement between larger and more massive objects. Also, macroscopic oscillators may be useful as quantum memories or as transducers of quantum information. ACMD researchers contributed to a PML experiment that demonstrated entanglement between two "drumhead" oscillators with masses of 70 pg each (see Figure 71). Their analysis showed that significant entanglement is present and directly observable, even without correcting for measurement inefficiency.

[1] S. Geller, S. Glancy, and E. Knill. Improving Quantum State Detection with Adaptive Sequential Observations. In preparation.

[2] R. Srinivas, S. C. Burd, H. M. Knaack, R. T. Sutherland, A. Kwiatkowski, S. Glancy, E. Knill, D. J. Wineland, D. Leibfried, A. C. Wilson, D. T. C. Allcock, and D. H. Slichter. High-Fidelity Laser-Free Entanglement and Addressing of Trapped Ion Qubits. In preparation.

[3] A. Avagyan, E. Knill, S. Glancy, and H. Vasconcelos. State Tomography with Photon Counting after a Beam Splitter. In preparation.

[4] I. Bezera, H. Vasconcelos, and S. Glancy. Estimation of Squeezing and Temperature from Fock Number Measurements. In preparation.

[5] S. Kotler, G. A. Peterson, E. Shojaee, F. Lecocq, K. Cicak, A. Kwiatkowski, S. Geller, S. Glancy, E. Knill, R. W. Simmonds, J. Aumentado, and J. D. Teufel. Tomography of Entangled Macroscopic Mechanical Objects. Preprint arXiv:2004.05515, 2020.

Correlated Noise in Quantum Devices

Yi-Kai Liu
Mohammad Hafezi (University of Maryland)
Kaixin Huang (University of Maryland)
Alireza Seif (University of Maryland)

Noise and decoherence are an important concern in the development of quantum devices for metrology and computation. While this topic has been studied for many years, it has recently gained new significance, as experimentalists start to build so-called NISQ (noisy intermediate-scale quantum) devices, which have tens to hundreds of qubits, with noise levels that are low but not insignificant. Two important aspects of this problem are the role of spatial correlations between qubits, which can become increasingly complicated when there are many qubits, and the role of non-Markovian memory effects, which are becoming more important as experimentalists build qubits with longer and longer coherence times.

We are developing various methods for modeling these kinds of effects and fitting these noise models to experimental data. First, we are developing a model of correlated dephasing noise on many qubits, where the correlations can be long-range, but are assumed to be sparse (in the sense that they are described by a sparse graph, which is not known a priori). This model is applicable to physical systems where two or more qubits can be coupled to the same environmental degree of freedom (e.g., superconducting qubits coupling to the same bosonic mode, due to spectral (frequency) crowding). In the past year, we have developed an improved method for fitting this model to experimentally accessible data, by preparing entangled W states, performing decay measurements with an adaptive rule for choosing how long to let the system evolve, and using compressed sensing techniques with rigorous recovery guarantees.

Second, we are developing more efficient methods for noise spectroscopy, for characterizing non-Markovian noise. In the past year, we have developed a general method for constructing random pulse sequences, with carefully controlled correlations between the pulses, which can be used to measure arbitrary linear functionals of the power spectral density of the noise. This method is relevant to physical systems such as superconducting qubits and nitrogen vacancy centers. It has several applications, such as characterizing noise spectra that consist of a few isolated peaks, estimating the parameters of noise spectra that have a Lorentzian shape, and learning effective representations of the environment (such as one-dimensional chain representations of a bosonic bath).

[1] A. Seif, M. Hafezi, and Y.-K. Liu. Compressed Sensing Measurement of Long-Range Correlated Noise. In preparation.

[2] K. Huang, A. Seif, M. Hafezi and Y.-K. Liu. Random Pulse Sequences for Qubit Noise Spectroscopy. In preparation.

Symmetry Breaking and Quantum Error Correction in Open Systems

Victor V. Albert
Alexey V. Gorshkov (NIST PML)
Simon Lieu (University of Maryland)
Ron Belyansky (University of Maryland)
Jeremy T. Young (University of Maryland)
Rex Lundgren (University of Maryland)

In order to realize the level of control required for quantum applications, one must carefully engineer the environment with which a quantum system interacts.

For example, in order to stabilize the subspace of an error-correcting code, one can engineer a Markovian master equation ("Lindbladian") whose steady-state subspace is exactly the space used for the encoding. Such a scheme has already been realized with conventional error-correcting codes tailored to photonic microwave cavities [1].

In a system whose evolution is governed by a Lindbladian, an initial state decays into a steady state in the limit of infinite time. In an error-correction setting (such as the example [1]), there is more than one possible steady state. The presence of multiple steady states is often associated with a symmetry, warranting an investigation into how such a symmetry could break and whether such symmetry breaking has any ramifications for the robustness of the system to noise. Such symmetry breaking mechanisms for Lindbladians are less well understood than those of closed quantum systems, in part because of their richer steady-state and symmetry structure. It turns out [2], for example, that a unitary symmetry in an open system can be imposed in either a "weak" or a "strong" way.

In this work, we characterize the possible Z_n symmetry breaking transitions for both weak and strong cases, making contact with the error correcting properties of the system. In the case of Z_2, a weak-symmetry-broken phase guarantees at most a classical bit steady-state structure, while a strong-symmetry-broken phase admits a partially protected steady-state qubit. Viewing the scheme [1] through the lens of strong-symmetry breaking, we show how to dynamically recover the logical information after any gap-preserving strong-symmetric error; such recovery becomes perfect exponentially quickly in the number of photons.

Our study forges a connection between driven-dissipative phase transitions and error correction. This work is published [3] and has been highlighted by the NIST Public Affairs Office.[34]

[1] Z. Leghtas, S. Touzard, I. M. Pop, A. Kou, B. Vlastakis, A. Petrenko, K. M. Sliwa, A. Narla, S. Shankar, M. J. Hatridge, M. Reagor, L. Frunzio, R. J. Schoelkopf, M. Mirrahimi, and M. H. Devoret. Confining the State of Light to a Quantum Manifold by Engineered Two-Photon Loss. *Science* **80** (2015), 347. DOI: 10.1126/science.aaa2085

[2] B. Buča and T. Prosen. A Note on Symmetry Reductions of the Lindblad Equation: Transport in Constrained Open Spin Chains. *New Journal of Physics* **14** (2012), 073007. DOI: 10.1088/1367-2630/14/7/073007

[3] S. Lieu, R. Belyansky, J. T. Young, R. Lundgren, V. V. Albert, and A. V. Gorshkov. Symmetry Breaking and Error Correction in Open Quantum Systems. *Physical Review Letters* **125** (2020), 240405. DOI: 10.1103/PhysRevLett.125.240405

[34] http://www.nist.gov/news-events/news/2020/12/error-prone-quantum-bits-could-correct-themselves-nist-physicists-show

Bipartite Energy-Time Uncertainty Relation for Quantum Metrology with Noise

Victor V. Albert
Philippe Faist (Freie Universität Berlin)
Mischa P. Woods (ETH Zurich)
Joseph M. Renes (ETH Zurich)
Jens Eisert (Freie Universität Berlin)
John Preskill (Caltech)

Quantum mechanics is full of tradeoffs, the most famous one being the Heisenberg uncertainty relation between the precision with which one can measure an object's position and momentum. The more precise the measurement of one of these "conjugate" variables, the less information one is able to extract about the other.

Another tradeoff often encountered in quantum error correction is the tradeoff between the effect of an observer and the uncontrolled environment on a quantum system. Typically, the less information the environment is able to extract from a quantum system, the more will be preserved for a later observer. One can think of the environment as a malevolent observer, whose clandestine probing of the system obscures the useful quantum information stored within the system. In this work, we determine a similar tradeoff (between the observer and the environment) in the sensitivity to determining an unknown parameter (such as time) in the evolution of a quantum state.

We study the local sensitivity of a quantum probe to measuring time, after the application of a given noise channel. We show that the decrease in sensitivity due to the noise is equal to the sensitivity that the environment gains with respect to the energy of the probe. We obtain necessary and sufficient conditions for when the probe does not suffer any sensitivity loss; these conditions are analogous to, but weaker than, the Knill-Laflamme quantum error correction conditions [1]. New upper bounds on the sensitivity of the noisy probe are obtained via our uncertainty relation, by applying known sensitivity lower bounds on the environment's system. This uncertainty relation asserts a general trade-off between the sensitivities that two parties can achieve for any two respective parameters of a single quantum system.

[1] E. Knill and R. Laflamme. A Theory of Quantum Error Correcting Codes. *Physical Review A* **55** (1997), 900.

Generalizing the Quantum Ising Model in Search of New Topological Defects

Victor V. Albert
David Aasen (Microsoft & UC Santa Barbara)
Wenqing Xu (University of Maryland)
Wenjie Ji (UC Santa Barbara)
John Preskill (Caltech)
Jason Alicea (Caltech)

Topological quantum computation, in contrast to the many other available blueprints for a quantum computer, holds the promise of computation in an inherently robust fashion. A paradigmatic example of such physics in 1D systems is the Ising quantum spin chain. With its edge, spin, and fermionic contexts, the Ising model yields a web of connections that brings topological physics to a broad array of seemingly unrelated systems. When viewed as a fundamentally fermionic system (via the Jordan-Wigner transformation), the model admits a topological phase hosting unpaired Majorana zero modes [1] that underlie non-Abelian statistics and provide the foundation for topological qubits resilient to local noise. When viewed as a standalone effective theory for edges of the toric code [2], the Majorana zero modes correspond either to anyonic-excitation operators (a.k.a. ribbons) lying on the edge or "defects" between different edge Hamiltonians. Continuum versions of the model, the (free) boson and its fermionic equivalent, the (noninteracting) Tomonaga-Luttinger liquid, have shined light on how to construct Majorana modes out of emergent degrees of freedom in electronic systems.

The most straightforward extension of the toric code (associated with spins taking values in the order-two group Z_2) is its Z_N version, which the Ising-like clock model has analogously provided connections between the toric code and various electronic systems that have witnessed encouraging recent experimental progress. Such connections, however, are less clear for the more intricate "nonabelian" generalizations of the toric code, quantum double models [2], defined on lattices whose spins take values in a nonabelian finite group G. Such systems house excitations described by nonabelian anyons, whose fusion can yield several outcomes (as opposed to just one in the abelian case). Information can be encoded in a robust fashion directly into superpositions of the fusion outcomes, and universal computation can be implemented with potentially less overhead than conventional schemes (e.g., without magic-state distillation). In contrast to the case of the conventional toric code, a standalone 1D effective theory describing the edge of a general quantum double model and its continuum description have yet to be examined.

Recently, Munk ,Rasmussen, and Burrello [3] introduced a G-site spin chain called the "flux ladder" and developed a corresponding generalization of the Jordan-Wigner mapping. However, direct connections of the flux ladder and the generalized mapping to quantum doubles remained unclear. In this work, we fill in this gap by deriving a standalone Ising-like model for a general quantum double edge and showing that this model is a lightly generalized flux ladder.

Armed with this model, we map out a web of connections for quantum double models, generalizing the existing web for the Z_N toric code. In a mapping distinct from that of [1], we develop Jordan-Wigner-like mode operators that are directly related to anionic ribbon operators of the quantum double model. We obtain a continuum description of the general flux ladder and recast a particular case in terms of fermions with the help of non-abelian bosonization. While our extensions arise from viewing the Z_N case, through the mathematical lens of representation theory the resulting physical connections to established systems should nevertheless brighten the prospects of eventual realization of quantum-double topological phases. Our paper on this topic is currently in preparation.

[1] A. Y. Kitaev. Unpaired Majorana Fermions in Quantum Wires. *Physics-Uspekhi* **44** (2001), 131. DOI: 10.1070/1063-7869/44/10S/S29.

[2] A. Yu. Kitaev. Fault-Tolerant Quantum Computation by Anyons. *Annals of Physics* **303** (2003), 2. DOI: 10.1016/S0003-4916(02)00018-0.

[3] M. I. K. Munk, A. Rasmussen, and M. Burrello. Dyonic Zero-Energy Modes. *Physical Review B* **98** (2018), 245135. DOI: 10.1103/PhysRevB.98.245135

Utilizing Configuration Space for Continuous-Variable Quantum Technologies

Victor V. Albert

Instead of building devices out of two-level components (qubits), the minimal component of conventional continuous-variable (CV) systems is the phase space associated with a canonical pair of continuous variables — position and momentum for a mechanical oscillator or quadrature components for an electromagnetic field mode. These systems are actively studied due to being both the key ingredients of four viable quantum technologies — microwave resonators, optical fiber modes, motional states of trapped ions, and nano-acoustic resonators.

CV states of defined position (or momentum) are labeled by points in the configuration space of a particle moving on a line. Most of this configuration space has remained unutilized in quantum applications, due to the

high energy cost required to pin a particle to a precise "position." However, with improved control, utilizing more of the configuration space harbors benefits for protecting quantum information and simulating quantum matter. For example, using the higher excitation subspace of a single system may require fewer resources than controlling a shallow collective low-energy subspace of a large number of copies of the system. In other words, adding an ancillary qubit to a given qubit requires control of the ancilla's corresponding physical system; instead, one can utilize the third or fourth excited states of the system whose ground and excited states make up the original qubit (see, e.g., the experiment [1]). As another example, by leveraging the continuous nature of the oscillator's configuration space, CV error-correcting codes can provide more information about errors and thus improve fault-tolerance thresholds (e.g., [2]). I would like to develop applications for CV platforms which are cognizant of the advantages that the oscillator's large configuration space can bring.

With control of quantum systems steadily improving, I believe it is also important to identify advantages in other physical configuration spaces. The inherent scalability and long-range dipole interactions of molecules make them a viable candidate for quantum information processing. However, as with the oscillator, we have to consider how to robustly encode information in order for this paradigm to fully mature. My collaborators and I have formulated molecular codes, protecting against small local shifts in both the molecule's (continuous) orientation and (discrete) angular momentum [3]. We have laid the groundwork for how to encode, process, and read out quantum information with asymmetric molecules and, more generally, systems whose configuration spaces are associated with a finite or compact group. I plan to extend these constructions to homogeneous spaces to develop codes for molecules with symmetries and field-theoretic configuration spaces.

[1] S. Rosenblum, P. Reinhold, M. Mirrahimi, L. Jiang, L. Frunzio, and R. J. Schoelkopf. Fault-Tolerant Detection of a Quantum Error. *Science* **361**:6399 (2018), 266-270.

[2] K. Fukui, A. Tomita, A. Okamoto, and K. Fujii. High-Threshold Fault-Tolerant Quantum Computation with Analog Quantum Error Correction. *Physical Review X* **8** (2018), 021054.

[3] V. V. Albert, J. P. Covey, and J. Preskill. Robust Encoding of a Qubit in a Molecule. *Physical Review X* **10** (2020), 031050.

Towards Robust Autotuning of Noisy Quantum Dot Devices

Joshua Ziegler
Justyna P. Zwolak
Jacob M. Taylor (NIST PML)
Sandesh Kalantre (University of Maryland)
Thomas McJunkin (University of Wisconsin-Madison)
Mark A. Eriksson (University of Wisconsin-Madison)

Gate-defined quantum dots, in which electrons in a semiconductor are trapped in electric potential wells, are a promising platform for quantum computing due to their small device footprint and the possibility of operation at few-Kelvin temperatures. However, due to minor inconsistencies inherent to the fabrication process, initialization of these devices with potentials that properly trap electrons cannot be achieved deterministically. Currently, initialization is performed mostly manually through heuristic tuning, although some recent progress has been made towards automating elements of this process.

The best strategies for automating initialization rely on supervised machine learning (ML) methods to autonomously identify the state of the device as parameters are tuned and use an algorithmic or ML optimizer to tune to a desired state [1, 2]. However, training such ML models requires a dataset with labeled device states. There are two paradigms for training, each having its own advantages and limitations. The first uses simulated data that is unrealistically noiseless compared to real experimental data, making the trained models less robust when implemented in an actual experiment. The other paradigm relies on manual labeling of experimental data, which is laborious and not scalable, especially as gate-defined quantum dot devices change or grow in

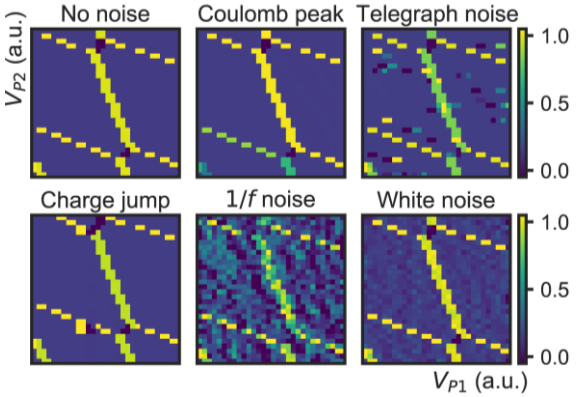


Figure 72. *Various types of noise added to a simulated charge stability diagram for a quantum dot device in a "double dot" state, with electrons trapped in two potential wells. The bright lines indicate a charge of electron occupation in the device. All noise magnitudes are the same as for the best accuracy quoted except for Coulomb peak which has increased effect for visibility.*

qubit numbers. Thus, current approaches force a choice between scalability and robustness.

In order to overcome these issues, we have expanded the capabilities of our quantum dot simulator to include various types of physics-based noise [3]. A noiseless charge stability diagram shows lines when an electron is added to the device's trapping potential landscape overlaid on a uniform background. Depending on the type of added noise, the appearance of simulated data changes as shown in Figure 72. We then trained ML models on that data and compared their performance to a model trained on noiseless data. For testing, we use a set of 232 manually labeled real experimental scans. Although our noise parameters and models are not yet fully optimized, our preliminary results show an accuracy improvement over our experimental dataset of about 29 %, from (55.6 ± 6.1) % when trained on raw simulated data to (84.4 ± 2.1) % when trained on simulated data with a combination of telegraph noise, 1/f noise, and white noise added.

With this approach, we expand the applicability of ML-based autotuning strategies to less ideal devices by making the ML models more robust. This is especially important when considering that future larger scale quantum dot devices will be fabricated in large foundries which may have difficulty matching the performance of devices fabricated in specialized academic facilities. More broadly, we show that making simulated data more physical can greatly improve the efficacy of ML models when deployed in a real lab environment, which may be a useful insight for other experiments combining ML and physics. An article reporting this work is currently being prepared [4].

[1] J. P. Zwolak, et al. Autotuning of Double-Dot Devices in Situ with Machine Learning. *Physical Review Applied* **13**:3 (2020), 034075.

[2] H. Moon et al. Machine Learning Enables Completely Automatic Tuning of a Quantum Device Faster Than Human Experts. *Nature Communications* **11**:1 (2020), 4161.

[3] J. P. Zwolak, S. S. Kalantre, X. Wu, S. Ragole, and J. M. Taylor. QFlow Lite Dataset: A Machine-learning Approach to the Charge States in Quantum Dot Experiments. *PLoS ONE* **13**:10 (2018), 1–17.

[4] J. Ziegler, S. S. Kalantre, T. McJunkin, M. A. Eriksson, J. M. Taylor, and J. P. Zwolak. Robust Automated Recognition of Noisy Quantum Dot States. In preparation.

Designing Polarization-Entangled Photon-Pair Sources

Paulina Kuo
Varun Verma (NIST PML)
Sae Woo Nam (NIST PML)

Entangled photon-pairs are critical parts of quantum networks. The photon pairs can be used to distribute entanglement across the network, thereby enabling quantum communications, distributed quantum computing and quantum-enhanced sensing. Photon polarization is a common way to encode the quantum information. In our work, we developed tools to design polarization-entangled photon pair sources.

To generate a Bell state such as $(|0\rangle + |1\rangle)/\sqrt{2}$ using spontaneous parametric down-conversion (SPDC), there must be two possible down-conversion paths that correspond to the two terms of the Bell state. The two down-conversion paths can be achieved by having two separate SPDC crystals. Alternatively, by using a specially designed aperiodically poled lithium niobate (aPPLN) crystal, we can simultaneously generate photon pairs with $|H_s V_i\rangle$ and $|V_s H_i\rangle$ polarizations (where the subscripts s and i represent signal and idler photons, and H and V are horizontal and vertical polarizations).

We modeled down-conversion to understand how aPPLN crystals compared to conventional SPDC crystals [1]. Figure 73 shows the results of modeling SPDC producing $|H_s V_i\rangle$ and $|V_s H_i\rangle$ states with a 776 nm pump and signal and idler at 1535.5 nm and 1571.5 nm, respectively. The figures show normalized SPDC intensity of the H polarization at different positions inside the crystal. Figure 73a plots SPDC in two consecutive crystals where the H polarization is first generated at 1535.5 nm (the signal) and then at 1571.5 nm (the idler). Figure 73b and Figure 73c show SPDC in aPPLN crystals where the $|H_s V_i\rangle$ and $|V_s H_i\rangle$ are simultaneously generated in a distributed fashion throughout the crystal. Figure 72d plots SPDC in an interlaced biperiodic structure [2] that effectively toggles between generating H-polarized photons at the signal and idler wavelengths. In the structures illustrated in Figure 73b through Figure 73d, spectral side-lobes can be seen, which are due to the presence of higher-order Fourier components in the periodic designs. One interesting discovery from this modeling is that the SPDC spectral bandwidth from the aPPLN and the interlaced biperiodic structures are half as narrow as the bandwidth when using two consecutive crystals.

We fabricated the aPPLN crystals and used them to generate entangled-photon pairs. We experimentally observed excellent polarization-entanglement visibility using these crystals [1]. This work represents a promising method to develop new entangled photon-pair sources.

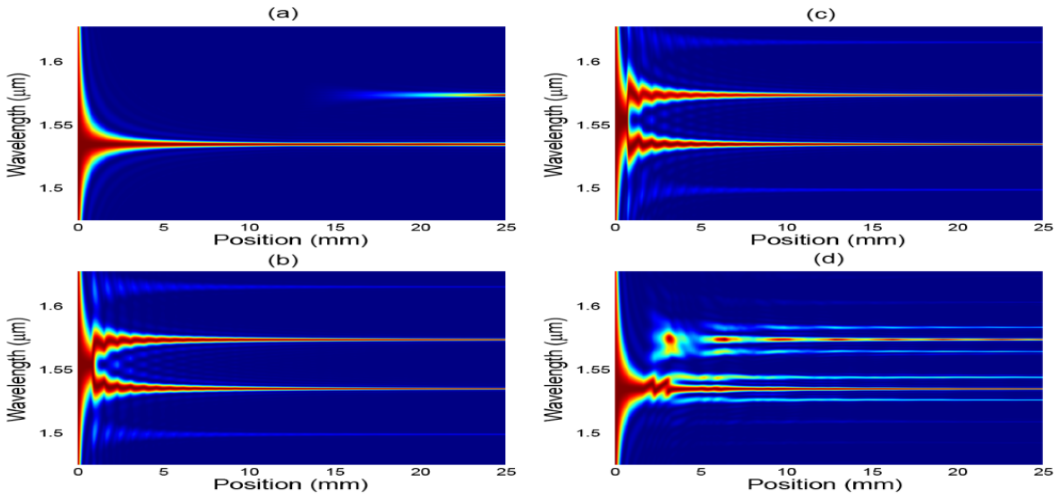


Figure 73. *Normalized SPDC intensity at different positions inside a crystal for (a) two consecutive crystals, (b) aPPLN crystal with dual-periodic-poling, (c) aPPLN crystal using phase-modulated and (d) an interlaced biperiodic structure with N=175. All crystals have the same total length of 25 mm.*

[1] P. S. Kuo, V. B. Verma, and S. W. Nam. Demonstration of A Polarization-Entangled Photon-Pair Source Based on Phase-Modulated PPLN. *OSA Continuum* **3**:2 (2020), 1260. DOI: 10.1364/OSAC.387449

[2] H. Herrmann, X. Yang, A. Thomas, A. Poppe, W. Sohler, and C. Silberhorn. Post-Selection Free, Integrated Optical Source of Non-Degenerate, Polarization Entangled Photon Pairs. *Optics Express* **21**:23 (2013), 27981.

Quantum Communications and Networking R&D

Oliver Slattery
Lijun Ma
Xiao Tang
Thomas Gerrits
Anouar Rahmouni
Summit Bhushan

http://www.nist.gov/quantum

Our research covers two areas:

- The creation, transmission, interfacing, storage, processing and measurement of optical qubits, the quantum states of single photons. We build and study quantum devices, such as entangled-photon sources, single-photon detectors, optical quantum memory and quantum interfaces. Our progress in these areas have been detailed in previous project summaries.
- The development of quantum network metrology tools and methods with which the performance of new and existing quantum devices and systems can be studied in a real-life network environment. These tools and methods will form part of a testbed to enhance the development of best-practices and protocols for quantum networks. We will use this year's project summary to introduce the motivation, our planned tasks and our recent progress in this newer area.

As the development of component quantum devices in the laboratories continues, more attention is being paid to testing these components in connected systems such as quantum networks. Characterization of photonic-based quantum components as well as validation and benchmarking their operation in a real-world setting is essential for ironing out issues with new technologies that strive to successfully make it into the emerging quantum network market. A quantum network will consist of many operational quantum "nodes," connected by optical fiber, which may receive, store, reroute, process and send quantum information carried on photons using a variety of quantum components (Figure 74). Measuring and enhancing the performance of these components, the quantum nodes and the wider quantum network, is the key motivation for our quantum network metrology efforts.

The main tasks for this project will be to identify the key performance parameters in quantum networks; to develop versatile tools and measurement procedures for quantum networks and quantum network components; and to collaborate within and beyond NIST to support the implementation of quantum network emulation, control, management and operation (Figure 75).

High-performance and accurately calibrated single-photon sources and detectors will be essential measurement tools of our quantum metrology efforts. During this year, we have begun the development of a cryogenic-based superconducting nanowire single-photon detector system that will be very accurately calibrated using existing NIST calibration methods. In addition,

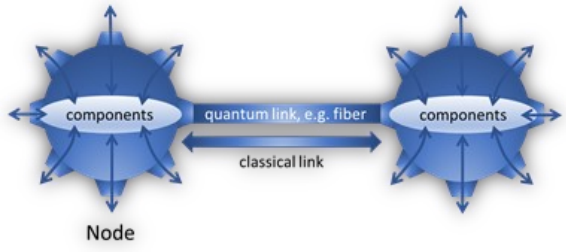


Figure 74. *A quantum network will consist of operational quantum nodes connected by fiber.*

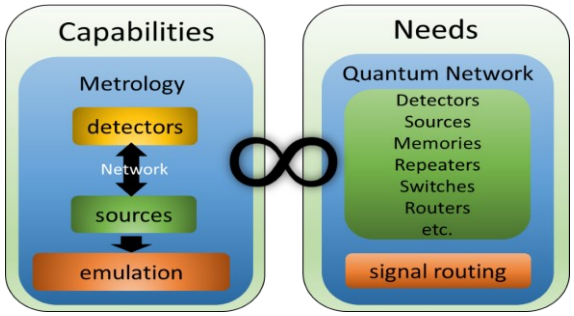


Figure 75. *High-performance and accurately calibrated single-photon sources and detectors will be essential tools to study the performances of quantum network components. We are currently building these tools.*

with the support of ITL's Building the Future (BTF) program, we have begun the development of an efficient, well-defined and well calibrated entangled photon source. Both of these components will be further developed to be portable so as to allow our metrology methods to be used both in the lab and in the field.

Once we have a source-detector system, we can start characterizing other network components, such as switches, fibers, beam splitters, transducers, memories, repeaters, and so on. From these characterizations we would like to be able to predict how a quantum state sent through a component could get altered or could decohere such that we can emulate these components on a large-scale network. This emulation will be of crucial importance for the pre-testing and development of quantum network components, and accurate measurements are the basis for accurate emulation predictions.

Joint Center for Quantum Information and Computer Science

Victor Albert
Matthew Coudron
Yi-Kai Liu
Carl Miller (NIST ITL)
Jacob Taylor (NIST PML)
Andrew Childs (University of Maryland)

http://quics.umd.edu/

Established in October 2014, the Joint Center for Quantum Information and Computer Science (QuICS) is a cooperative venture of NIST and the University of Maryland (UMD) to promote basic research in understanding how quantum systems can be effectively used to store, transport and process information. QuICS brings together researchers from the University of Maryland Institute for Advanced Computer Studies (UMIACS) and the UMD Department of Physics and Computer Science with NIST's Information Technology and Physical Measurement Laboratories, together with postdocs, students and a host of visiting scientists.

QuICS has quickly established itself as a premier center for research in quantum information science. Twelve Fellows, one Adjunct Fellow, 19 postdocs, and 38 graduate students are currently associated with the center. Two new Fellows are expected to join the Center in FY 2021. This year Matthew Coudron and Victor Albert of ACMD joined Yi-Kai Liu of ACMD and Carl Miller of the ITL Computer Security Division as QuICS Fellows, each holding adjunct appointments at the University where they mentor students and postdocs. Yi-Kai Liu is Co-Director of the center along with Andrew Childs of the UMD Computer Science Department.

The Center is very productive. In CY 2020 some 120 research papers were produced by those associated with the center. These are often published in top venues. For example, three papers by QuICS researchers were accepted to the 2020 Symposium on the Theory of Computing (STOC), which is one of the two top conferences in theoretical computer science. In January 2020, QuICS held a one-day symposium in honor of its 5th anniversary, which featured luminaries such as John Preskill (Caltech), Barbara Terhal (TU Delft/QuTech), and Eddie Farhi (MIT/Google) as keynote speakers.

QuICS' success has resulted in it outgrowing its original space in the Atlantic Building on the College Park Campus. Happily, QuICS has obtained new adjacent space which has doubled its footprint there, including 15 new offices and an expanded seminar room. The space awaits a return to normal operation when the campus re-opens to all.

Figure 76. *Panelists at the QuICS 5 Year Anniversary Symposium (from left to right): Ronald Boisvert (ACMD), Robert Cunningham (LPS), Eddie Farhi (MIT/Google), Laurie Locascio (UMD), John Preskill (Caltech), Barbara Terhal (TU Delft/QuTech), Jake Taylor (NIST PML). Photo courtesy of the University of Maryland.*

Foundations of Measurement Science for Information Systems

ITL assumes primary responsibility within NIST for the development of measurement science infrastructure and related standards for IT and its applications. ACMD develops the mathematical foundations for such work. This can be very challenging. For example, many large-scale information-centric systems can be characterized as an interconnection of many independently operating components (e.g., software systems, communication networks, the power grid, transportation systems, financial systems). A looming new example of importance to NIST is the Internet of Things. Exactly how the structure of such large-scale interconnected systems and the local dynamics of its components leads to system-level behavior is only weakly understood. This inability to predict the systemic risk inherent in system design leaves us open to unrealized potential to improve systems or to avoid potentially devastating failures. Characterizing complex systems and their security and reliability properties remains a challenging measurement science problem.

Towards Actionable Cybersecurity Risk Metrics in a Hyperconnected World

Brian Cloteaux,
Vladimir Marbukh
Debasis Mitra

Real systems contain potential vulnerabilities which can be exploited by adversaries in attempt to disrupt system operations. Emergence of large-scale critical mission infrastructures makes quantitative evaluation of the overall system security risk not only exceedingly urgent but also very challenging due to large number of potential vulnerabilities and non-linear contribution of individual vulnerabilities to the overall system risk. Conventional measures of the security risk by the expected economic loss due to successful exploits of potential vulnerabilities is often inadequate due to disregard for the tail risk, inherent inaccuracy of estimates of low probabilities, and reliance on the attacker(s) model.

The recently proposed measure of cybersecurity risk, Cybersecurity Value at Risk (CyVaR) [1], which is based on a Value at Risk (VaR) measure of financial risk, accounts for the tail risk. However, CyVaR still suffers from its own problems, e.g., is not a coherent risk measure. Following the recent trend of replacing VaR with Entropic VaR (EVaR) as a measure of financial risk [2], we suggest replacing CyVaR with CyEVaR, which is a robust and coherent risk measure [3]. CyEVaR, which has natural game-theoretic interpretation as a game against a boundedly rational adversary, interpolates between Bayesian and purely adversarial attack models. Figure 77 shows CyEVaR as a function of the attack severity measured by the number of successfully exploited vulnerabilities s. As s increases from the expected number $\tilde{s}$ to the total number of vulnerabilities N, CyEVaR grows from the average loss $\tilde{L}$ to the maximum possible loss $\hat{L}$.

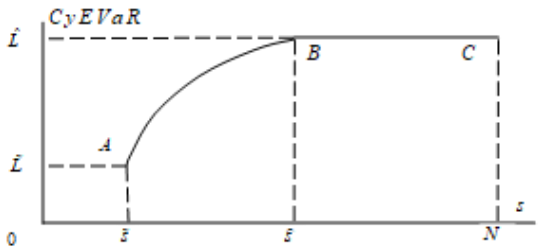


Figure 77. *Cyber Entropic Value at Risk vs. attack severity.*

Our current work includes generalization of the Gordon-Loeb model of optimal investments in cybersecurity [4] in order to make this model applicable to systems/enterprises described by Bayesian Attack Graphs. This generalized model identifies the optimal mixture of investments in mitigation of various system vulnerabilities. To this end, we created various attack scenarios, e.g., a randomized scenario characterized by average loss and a strategic scenario characterized by maximum loss. Separate from our analytic investigations, we have explored computational approaches to optimal investments in instances of complex Bayesian Attack Graphs with multiple entry points for the attacker and multiple system breach nodes.

We are also extending our research on network evolution driven by competing incentives for higher connectivity on the one hand and infection avoidance on the other hand [5]. These extensions include networks evolving by rewiring, infection avoidance under partial information, etc.

Our plans for future investigations include extending computational approaches for complex graph structures, developing models for cycles of attacker-defender action-reaction interactions, developing machine learning techniques for the attacker's evolving strategy in interaction cycles, and extending interaction models from single vulnerability to network models.

Our ultimate goal is quantification of the existing qualitative cybersecurity risk mitigation recommendations, e.g., NIST Cybersecurity Framework [6]. Achieving this goal will require significant efforts in developing computationally effective yet accurate

estimates of the Security Risk Reduction Return on Investment (SRR-RoI). In our future research we will employ computational techniques of statistical physics to approximate CyEVaR, and employ CyEVaR decomposition techniques, including Shapley value, to approximate SRR-RoI.

[1] World Economic Forum. Partnering for Cyber Resilience Towards the Quantification of Cyber Threats. 2015. URL: http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf

[2] A. Ahmadi-Javid. Entropic Value-At-Risk: A New Coherent Risk Measure. *Journal of Optimization Theory and Applications* **155**:3 (2012), 1105–1123.

[3] V. Marbukh. Towards Robust Security Risk Metrics for Networked Systems: Work in Progress. In *2021 IFIP/IEEE International Symposium on Integrated Network Management* (IEEE IM 2021), to appear.

[4] L. Gordon and M. Loeb. The Economics of Information Security Investment. *ACM Transactions on Information and System Security* **5**:4 (2002), 438–457.

[5] B. Cloteaux and V. Marbukh. SIS Contagion Avoidance in a Network Growing by a Preferential Attachment: Work in Progress. In *GRADES-NDA'19: Proceedings of the 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics*, June 2019, 11.

[6] NIST Cybersecurity Framework. URL: https://www.nist.gov/cyberframework

Cyber Risks of Complex Systems and Optimal Cybersecurity Investments

Richard J. La
Van Sy Mai (NIST ITL)
Abdella Battou (NIST ITL)

Today, many modern engineered systems, including information and communication networks and power systems, comprise many interdependent systems. For uninterrupted delivery of their services, the comprising systems must work together and oftentimes support each other. Unfortunately, this interdependence among comprising systems also introduces a source of vulnerability in that it is possible for a local failure or infection of a system by malware to spread to other systems, potentially compromising the integrity of the overall system. For instance, an outage in one part of an electrical grid, such as a transmission line failure, can trigger cascading failures and lead to a large-scale blackout, affecting millions of people (e.g., Northeast blackout of 2003 and India blackouts of 2012). Analogously, in social networks, contagious diseases often spread from infected individuals to other vulnerable individuals through contacts or physical proximity.

From this viewpoint, it is clear that the underlying networks that govern the interdependence among systems have large impact on dynamics of the spread of failures or malware infections. Similarly, the topology and contact frequencies among individuals in social network significantly influence the spread of diseases in societies. Thus, any sound investments in (cyber)security of complex systems or the control of epidemics ought to consider the interdependence in the systems and social contacts in order to maximize potential benefits from the investments.

With this in mind, we studied two related problems. First, we investigated the problem of designing computationally efficient methods for identifying nodes that are most likely to trigger cascading failures in complex systems. Such nodes serve as good candidates for "hardening" to enhance the overall robustness of complex systems to failures and infections. Building upon our earlier eigenvector-based approach [1], we proposed a new computationally efficient method for identifying vulnerable nodes (i.e., an optimization-based approach [2]) which, even in the absence of sufficient historical data or detailed simulation, can find many, if not most, of vulnerable nodes that are susceptible to triggering cascading failures in the system. This new method does not suffer from a shortcoming of the eigenvector-based approach that can lead to degraded performance when the Perron-Frobenius eigenvalue of the associated Hashimoto matrix deviates significantly from one.

Second, we studied a closely related problem of finding optimal cybersecurity investments for hardening vulnerable nodes, which will minimize the long-run average costs of systems that account for both security investments and recovery/repair costs ensuing infections or failures, which we call infection costs. To this end, we first considered a scenario where malicious actors launch external or primary attacks. In the second case, we assume that there are no primary attacks and examine the steady state, starting with an initial state where some nodes are infected. The goal of studying the second scenario is to get additional insights into scenarios where the primary attacks occur infrequently. It turns out that, even in the absence of primary attacks, infections may persist due to secondary attacks by infected systems on other systems and, as a result, may not be able to attain the infection-free steady state.

We formulated the problem of determining the optimal security investments that minimize the average costs as an optimization problem [3, 4]. Unfortunately, this optimization problem is nonconvex and cannot be solved easily. In order to gauge the quality of a feasible solution, we obtained both a lower bound and an upper bound on the optimal value of our problem. A lower bound can be acquired using one of two different convex relaxations of the original problem we proposed. For these convex relaxations, we also derived a sufficient

condition under which the solution of the convex relaxation solves the original nonconvex optimization problem. An upper bound on the optimal value can be obtained using an algorithm that finds a local minimizer. To this end, we proposed two methods: a reduced gradient method (RGM) and sequential convex programming, both of which produce a local minimizer. Together, these offer a bound on the optimality gap.

Numerical studies show that the computational requirements for the proposed methods are light to modest even for large systems, except for one method, which requires the calculation of an inverse matrix [4]. They suggest that, in almost all cases that we considered, the gap between the lower bound on the optimal value and the cost achieved by our solutions is small; in fact, in most cases, the gap is less than 2-3 % with the gap being less than 0.3 % in many cases. In addition, when the infection costs are large, which are likely true in many practical scenarios, the sufficient condition for the convex relaxations to be exact holds, and we obtain optimal points by solving the convex relaxations. Finally, the RGM is computationally most efficient (with the computational time being less than two seconds in all considered cases and less than 0.1 seconds in most cases) and the quality of solutions is on par with that of other methods. This suggests that the RGM may offer a good practical solution for our problem.

[1] R. J. La. Identifying Vulnerable Nodes to Cascading Failures: Centrality to the Rescue. In *Complex Networks and Their Applications VII. COMPLEX NETWORKS 2018.* Studies in Computational Intelligence **812**, Springer, 2019.

[2] R.J. La. Identifying Vulnerable Nodes to Cascading Failures: Optimization-Based Approach. In *Complex Networks and Their Applications VIII. COMPLEX NETWORKS 2019.* Studies in Computational Intelligence **881**, Springer, 2020.

[3] V.-S. Mai, R.J. La, and A. Battou. Optimal Cybersecurity Investments for SIS Model. *Complex Networks and Their Applications VIII. COMPLEX NETWORKS 2019.* Studies in Computational Intelligence **881**, Springer, 2020.

[4] V.-S. Mai, R.J. La, and A. Battou. Optimal Cybersecurity Investments in Large Networks Using SIS Model: Algorithm Design. In review.

Algorithms for Identifying Important Network Nodes for Communication and Spread

Fern Y. Hunt
Roldan Pozo

The identification of nodes in a network that enable the fastest spread of information is an important if not fundamental problem in network control and design. It is applicable to the optimal placement of sensors, the design of secure networks, and the problem of control when network resources are limited. Our approach to this problem has its origins in models of opinion dynamics and the spread of innovation in social networks. The mode of communication between nodes is described by simple models of random or deterministic propagation of information from a node to its neighbors. During the past few years, we have made progress in understanding the structural requirements for sets of nodes for effective spread in networks and have developed scalable algorithms for constructing these sets in real world networks.

We consider a discrete time model of information spread (represented by a variable assigned to each node) in a network with a set of nodes V and a subset $A \subseteq V$ of k nodes representing leaders or stubborn agents that are initially assigned a single value. Propagation occurs by iterated averaging or diffusion defined by a stochastic matrix P. All node values will eventually converge to the single value at a speed determined by the sub-stochastic matrix P~A, the matrix P restricted to the complement of A. An effective spreader in this situation is then a set of nodes for which convergence to this single value is fastest, i.e., the set A for which the Perron-Frobenius eigenvalue of P~A is largest. Using a classical result of Markov chain theory, the problem can be recast in terms of finding the set A of cardinality k that minimizes the mean first hitting time, i.e., the expected time a random walker reaches the target set A for the first time.

Previously we considered a polynomial time algorithm for finding an approximation to the optimal set. It is an extension of the classic greedy algorithm and it begins with a class of optimal and near optimal starter sets of smaller cardinalities rather than the conventional choice of a best singleton set. Direct comparison of the algorithm results with the actual optimal solution and lower bounds on the performance ratio can be obtained because F is a supermodular set function [2]. However, for large complex networks commonly encountered in applications, another approach is needed.

Recently, we developed a set of fast heuristics that work well on graphs with large hubs, a common feature of complex networks. When the desired set cardinality is k, subsets of hub vertices are rapidly screened to produce candidate sets. Each set consists of k nodes whose first (or higher order) neighborhoods have minimal

overlap. After further screening, the offered approximation is selected by ranking the results of a Monte Carlo calculation of the optimal set F for each candidate. This process allows us to find near optimal and optimal spreaders in networks with millions of nodes and dozens of millions of edges in less than a few seconds on a typical laptop. After conducting tests on real world graphs from diverse application areas including molecular biology, traffic control and social networks, we hypothesize that the method is most effective in terms of speed and quality of offered solutions when it is used on graphs with a large ratio of maximal degree to average degree.

This past year, we continued our investigation of the accuracy of this procedure. After realizing that the resulting offered set was an approximate solution of a discrete stochastic optimization problem, we established sufficient conditions that imply that it is also an approximate solution of the original problem. The first step was to establish the accuracy of the Monte Carlo calculation of F. The fact that the first hitting time to a set A has a distribution with exponential tails means that a sample average of simulated hitting times produces a consistent estimate of F in the limit of large sample size i.e. number of simulations.

Establishing the degree of optimality of any offered solution is very difficult since supermodularity cannot be used and the size of the graphs are so large. However, the methods we use make it possible to rapidly sample the distribution of possible F values. We suppose the screening and ranking procedures produce candidate sets with F values that rank in the highest percentile of a distribution of such values over all subsets of fixed cardinality. Independent repetition of the heuristic calculation enables us to produce an estimate of a fixed percentile value along with a confidence interval for that estimate. The latter follows from an application of Chebyshev's inequality. Note that the resulting interval contains both the offered solution value and the optimal value of the original problem. Even in the case of a large number of repetitions, this approach is promising because it takes very little time to perform a single execution. The results of our work are reported in [3].

[1] F. Hunt. *Using First Hitting Times to Maximize the Rate of Convergence to Consensus*. Preprint arXiv:1812.08881, 2018.

[2] F. Hunt. An Algorithm for Identifying Optimal Spreaders in a Random Walk Model of Network Communication. *Journal of Research of the NIST* **121** (2016), 180-195.

[3] F. Hunt and R. Pozo, Fast Methods for Identifying Effective Spreaders in Real Network. *Journal of Research of the NIST* **125** (2020), 125036. DOI: 10.6028/jres.125.036

Estimating Undetected/Hidden Infections in Epidemic Processes

Richard J. La
Van Sy Mai (NIST ITL)

The spread of malware and viruses in computer networks and information systems, contagious diseases, rumors or new product information is often studied using epidemic processes. Many, if not most, of these studies assume that the infected nodes and, in many cases, the sequence of infections, are made known so that appropriate parameters associated with the epidemic processes, such as reproduction numbers, can be estimated from the available data. In some cases, however, such information may not always be available or cannot be obtained easily. For example, when devices on a network, such as computers and servers, with inadequate security measures are infected, the users and network managers may not be aware that their devices are compromised for some time. This concern is growing with the emergence of artificial intelligence-powered malware [1, 2]. Similarly, a contagious disease can cause widely varying symptoms and some infected individuals may exhibit no obvious outward symptoms that can be easily identified.

Devising appropriate and effective cybersecurity measures or public health policies requires accurate information. This task would be more challenging when it is difficult to obtain a clear picture of how aggressively or quickly computer malware/viruses or contagious diseases spread in a network or society; missing a large number of infections in reported cases at an early stage of an outbreak of a computer virus or a deadly pandemic would likely lead to an unfortunate and potentially costly underestimation of the severity of the problem, which could have been mitigated by timely countermeasures and policies. For this reason, it is important to quickly determine in an early stage of an epidemic process whether the numbers of reported cases correctly represent the true numbers of total infections or whether they are missing many undetected cases, so that cybersecurity or healthcare experts can introduce suitable countermeasures and remedies.

To this end, we proposed a new approach to determining whether the reported numbers of infection cases are accurate or whether there are many undetected, and hence unreported cases [3]. Our approach borrows insights from (statistical) model selection [4] and is based on a simple observation that infected devices or individuals that go undetected spread their infection differently than those that are detected shortly after suffering infection. Hence, our problem can be formulated as one of identifying the model that best describes or fits the available data. Equipped with this new method, we plan to evaluate the performance of our proposed approach using real data, both in cyberattacks and epidemiology.

[1] J. Menn. New Genre of Artificial Intelligence Programs Take Computer Hacking into Another Level. *Reuters*, August 8, 2018.

[2] B. Dickson. How Hackers Can Use AI to Hide Their Malware and Target You. *The Daily Dot*, August 27, 2018.

[3] R. J. La and V.-S. Mai. Determining the Presence of Undetected Infections in Epidemic Processes. In preparation.

[4] H. Linhart and W. Zucchini. *Model Selection*. John Wiley & Sons, 1986.

Statistical Change Detection for Network Anomalies

Assane Gueye
Anthony Kearsley
Pulkit Grover (Carnegie Mellon University)
Rosaline Su (Carnegie Mellon University)

Around the world, network traffic ebbs and flows, producing enormous amounts of observable information (e.g., data flow, queue times). There are instances where networks are affected by some outside interference and this, in turn, leads to changes in network performance and thus a change in observable operational statistics of the network. In this project we seek to study and characterize these changes. We hope to answer the following questions. Are these changes statistically significant and can they be used to indicate outside interference? If so, with what kind of certainty can we say interference has taken place and how quickly can we say it? Finally, can this lead to characterizing and classification of detected interferences?

Our approach is to examine time series representations of network statistics like queuing times, routers' load, and other time-dependent network characteristics. Initial data will be a straight-forward time-series histogram of a collection of statistical network data. Given a metric, a regression of this data can then be performed. However, it is clear that this regression will vary substantially as a function of selected metrics. Our current plans include a variety of metrics including the l_1, l_2, l_∞ vector norms, the X_2 metric and the Wasserstien metric, sometimes called the Earth Mover Distance (EMD). A stochastic regression of the time series can then decompose our time series into advection and diffusion where, presumably, the ambient noise present in every network should be identifiable.

A stochastic regression of our network observations yields a stochastic differential equation (SDE) comprised of a deterministic and stochastic component. If successful, this decomposition will isolate ambient noise in the network and allow more scrutiny of the underlying network behavior signal. Large variations in these two quantities can then be used to identify changes indicative of outside stimulation. Our hope is that if statistical quantities are well selected, malicious interference will become apparent from our regression. Roughly speaking, the regression should provide an estimate and statistical characteristics of an ordinarily running network and large variations from that baseline can be easily identified using this technique.

We hope to answer the following questions. Have the statistics changed in a way that indicates outside interference? If so, with what kind of certainty can we say this interference has taken place and when? Finally, is there a way to characterize the type of interference?

This project is with collaboration with the main Carnegie Mellon University campus in Pittsburgh and CMU-Africa (Kigali Campus).

[1] J. Fan, J. Jiang, C. Zhang, and Z. Zhou. Time Dependent Diffusion Models for Term Structure Dynamics. *Statistica Sinica* **13** (2003), 965–992.

[2] C. Manikopoulos and S. Papavassiliou. Network Intrusion and Fault Detection: A Statistical Anomaly Approach. *IEEE Communications Magazine* **40**:10 (2002), 76-82.

[3] V. Kotu and B. Deshpande. *Data Science*, Concepts and practice. Second Edition, Morgan Kaufmann, 2019, Chapter 13, 447-465.

[4] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Heidelberg, 1992.

[5] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *Int. Journal of Computer Vision* **40** (2000), 99-121.

Utility Proportional Resource Allocation in Virtualized Radio Access Networks

Vladimir Marbukh
Kamran Sayrafian
Behnam Rouzbehani (University of Lisbon, Portugal)

Virtualization platforms are responsible for allocation and aggregation of radio resources from different access technologies as well as the distribution of the total capacity among virtual network operators (VNOs). The radio resource management (RRM) employed by each VNO should comply with the requirements specified in the service level agreements (SLAs) of each user.

In our previous research [1, 2], we developed a joint admission control and resource management scheme based on proportionally fair rate allocation among different users. Although, all SLAs are satisfied in that scheme, users with vastly different QoS requirements might not necessary be treated fairly in terms of the allocated rates. This is especially the case when the

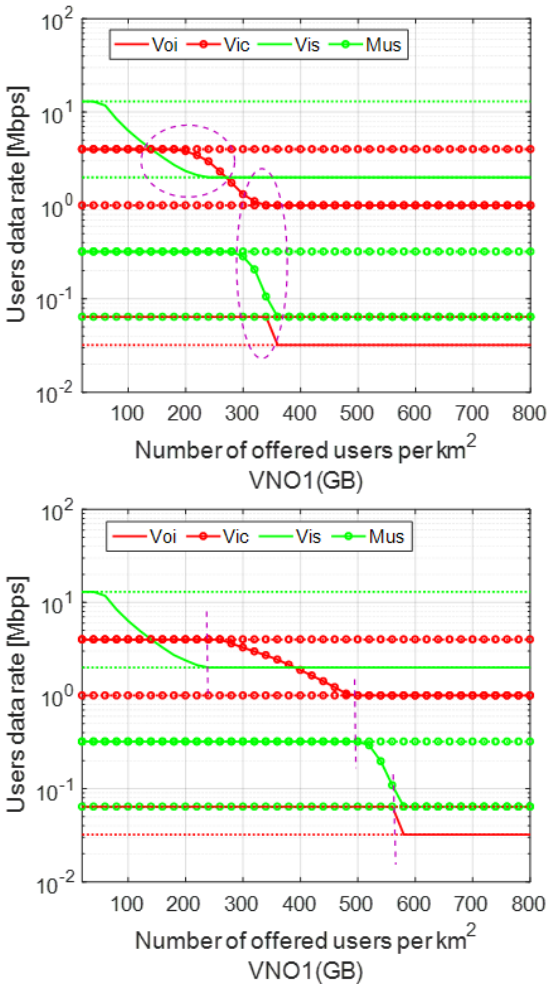


Figure 78. *User data rates for proportionally fair (top) and utility-proportional (bottom) rate allocations*

available capacities of the VNOs cannot support the maximum requested rates for all such users. This project attempts to overcome this weakness by replacing the proportional fairness strategy with a more general concept of utility-proportional fairness. We extended the proportionally fair rate allocation scheme in [1, 2] to a more general utility-proportional rate allocation. The intention is to address some of the observed shortcomings of proportionally fair allocation, such as giving advantage to users with low bandwidth requirements. Our proposed scheme maximizes the aggregate system utility using a two-stage distributed optimization on a fast and slow time scale and overcomes the scalability issues of the centralized RRM.

In general, under proportionally fair resource allocation, users with highest SLA rate requirements will encounter reduced assigned rates well before users with lower SLA rates. Under a "fair" rate allocation scheme, all users should experience rate reduction from their maximum assigned rates approximately around the same time. Figure 78 shows rate allocation in a VNO which provides four classes of services: voice (Voi), video calling (Vic), video streaming (Vis), and music streaming (Mus). Under proportionally fair resource allocation, as the number of users increase, Voi users that have the lowest rate SLA requirements maintain their maximum requested rate long after Vis user rates drop to their minimum requested rate. This problem is alleviated with a utility proportional rate allocation strategy.

The radio resource management scheme proposed in this research overcomes the shortcomings of the proportionally fair rate allocation by exploiting the general concept of utility proportional fairness [3]. One may expect that a customized SLA-based utility proportional fairness could lead to even better overall performance. We plan to further investigate this issue in future research. Viability of the time scale separation and convergence assumption in practical situations should also be studied. That will include mechanisms to mitigate performance loss in situations of comparable time scales in rate/capacity adaptation and users' arrivals/departures process. A requirement for this study is realistic models of users' arrival and departure processes. Finally, employing artificial intelligence techniques as a part of the network management could be a major focus of future research.

[1] B. Rouzbehani, V. Marbukh, K. Sayrafian, and L.M. Correia. Towards Cross-Layer Optimization of Virtualized Radio Access Networks. In *Proceedings of the 2019 IEEE European Conference on Networks and Communications* (EUCNC'19), Valencia, Spain, June 18-21, 2019.

[2] B. Rouzbehani, V. Marbukh, and K. Sayrafian. A Joint Admission Control & Resource Management Scheme for Virtualized Radio Access Networks. In *Proceedings of the 5th Annual IEEE Conference on Standards for Communications and Networking* (CSCN'2019), Granada, Spain, October 28-30, 2019.

[3] B. Rouzbehani, V. Marbukh, and K. Sayrafian. Utility Proportional Resource Allocation for Users with Diverse SLAs in Virtualized Radio Access Networks. In *Proceedings of the IEEE International Black Sea Conference on Communications and Networking* (BlackSeaCom'2020), Online, May 26-29, 2020.

An Immersive Interactive Tool to Select Sample Locations for Statistical Pathloss Modeling of Wireless Capsule Endoscopy

Wesley Griffin
Katjana Krhac
Kamran Sayrafian

Wireless capsule endoscopy (WCE) is a painless, effective, and novel diagnostic technology for inspecting the entire human gastrointestinal (GI) tract. A typical WCE

includes a miniature-sized camera that periodically transmits images as it passes through the GI tract. The images are captured by several on-body receivers for the duration of the procedure. Comprehensive study of radio wave propagation for ingestible electronics is quite challenging. The first step in this study involves an in-depth understanding of the wireless channel, and development of a statistical pathloss model for the communication link between the capsule and an on-body receiver.

To obtain this statistical pathloss model, a set of sample capsule location points inside the GI tract is needed. These capsule locations will define a set of simulation scenarios for radio wave propagation and signal measurements. The simple approach that comes to mind is to choose uniformly spaced locations along the GI tract centerline as sample points. However, since the GI tract is a complex intertwined path in 3D, uniformly spaced sample points could lead to scenarios where the capsule distances from the on-body receivers are heavily populated within certain ranges, while there are not enough sample representatives in other distance ranges. This will negatively impact the accuracy of the statistical analysis needed for the communication pathloss modeling. In [1], we proved the significance of a judicious selection of measurement points for statistical modeling using a simple liquid phantom platform. Our objective in that project was to obtain a set of spatially spread sample locations resulting in a more uniform distance histogram. This will ensure sufficient representation of the entire GI tract when statistical parameters of the pathloss model are calculated.

In order to enable location selection inside the GI tract, an interactive tool in the NIST CAVE[35] has been developed. The tool consists of several different modules with the following functions:

1. Identify points along the centerline of the GI tract that are a specific distance from any one receiver (small green spheres).
2. Select locations along the centerline of the GI tract (blue glyphs).
3. Show the distances between receivers and all selected points.

To initiate the selection process, points along the GI tract at specific distances from the receivers were computed. The distances were chosen to be from 5 cm to 25 cm away from a receiver in increments of 2 cm.

The NIST CAVE supports interactive probing and picking in the 3D scene. Using this capability, a user is able to select points along the GI tract centerline. However, we want to ensure that the point chosen has a specific distance to the receivers based on the precomputed points. Thus, we implemented a K-nearest-neighbor search using CGAL [2] that takes the user-chosen point and then finds the closest matching precomputed point. To ensure the user-selected points are uniformly distributed across the range of interest we display a histogram of distances between the on-body receivers and all selected capsule locations. Thus, the user can instantly see if the selected sample locations maintain the desired distribution and whether further adjustments (i.e., sample point addition/subtraction) are needed. The user will continue this process to judiciously select the final set. Figure 79 shows a user interaction at NIST CAVE during point selection process. A short video clip demonstrating the interaction can be found at [3].

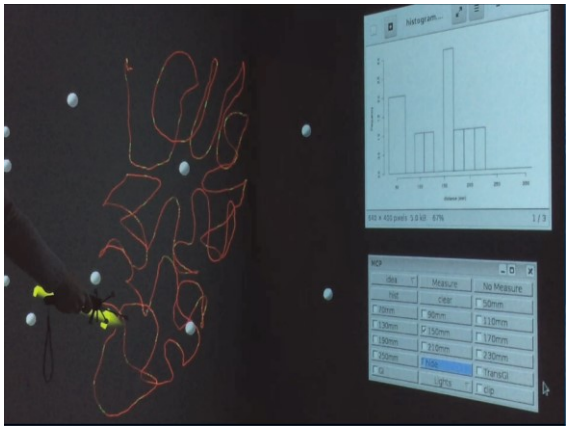


Figure 79. *The interactive tool used to select points (blue glyph) along inner GI tract (red line). The white spheres are the receiver locations. The inset histogram shows the distances from receivers to all selected points.*

While it is conceivable to formulate a complicated mathematical approach to solve for the capsule location problem, our interactive tool offered a visually attractive and relatively easy alternative to find the desired points. A statistical pathloss model of UWB frequencies is currently under development using these results.

[1] S. Perez-Simbor, K. Krhac, C. Garcia-Pardo, K. Sayrafian, D. Simunic, and N. Cardona. Impact of Measurement Points Distribution on the Parameters of UWB Implant Channel Model. In *Proceedings of the IEEE Conference on Standards for Communications* (CSCN), Paris, France, October 29-31, 2018.

[2] H. Tangelder and A. Fabri. dD spatial searching. In *CGAL User and Reference Manual*. CGAL Editorial Board, Edition 4.14, 2019.

[3] W. Griffin, K. Krhač, and K. Sayrafian. "Sample Location Selection for Statistical Pathloss Model of Wireless Capsule Endoscopy." Poster, IEEE VIS, Vancouver, Canada, October 19-25, 2019. URL: https://vimeo.com/361160006

[35] https://www.nist.gov/itl/math/visualization-body-area-networks

A Hybrid Machine Learning and Analytical Approach to Maximize the Output Power of Micro Energy Harvesters

Kamran Sayrafian
Masoud Roudneshin (Concordia University, Canada)
Amir G. Aghdam (Concordia University, Canada)

Recent advancements in micro-electronics have led to the development of miniature wearable sensors that can be used for a variety of health monitoring applications. These sensors are typically powered by small batteries which could require frequent recharge. Kinetic energy can be a reliable solution for auxiliary power generation, and hence holds promise for reducing the charging frequency of medical sensors. For cases of nonstationary vibrations (e.g., human body motion), a Coulomb force parametric generator (CFPG) architecture has been proposed as a promising solution to extract power from such movements [1]. Our objective in this project is to investigate mechanisms that can maximize the output power of a kinetic-based micro energy-harvester.

The mechanical power generated in such micro-harvesters (i.e., $P(t)$) is a function of an architectural parameter called the electrostatic holding force (F). Assuming that the electrostatic force can be adjusted based on the input acceleration every Δ seconds, our objective is to identify an optimal value of F for the next time interval according to the following optimization problem.

$$\underset{F_i}{\text{argmax}} \left[\frac{1}{\Delta} \times \sum_{t=t_i}^{\Delta+t_i} P(t) \right]$$

In general, we are searching for the optimal electrostatic force value which is a function of both the decision interval length Δ and the acceleration samples in the time interval $[t_0 + (i-1)\Delta, t_0 + i\Delta]$. This search can be implemented by using machine learning (ML) techniques such as Multi-Layer Perceptron (MLP), K-Nearest Neighbor (KNN), Support Vector Machine (SVM) or Random Forest Classifier in both frequency and time domains [2, 3]. Our initial investigation has shown that, in terms of accuracy, the MLP technique outperforms the other three approaches.

Machine learning algorithms are basically black-box approaches that estimate the value of the optimal electrostatic force in each measurement interval. In other words, they are mappings from the acceleration data to the optimal values of F without using a dynamic model of the CFPG. We have also developed and studied an adaptive approach for the estimation of the optimal electrostatic force by combining the above-mentioned ML techniques with a lower bound obtained through analysis of the underlying non-linear differential equation representing the system dynamics

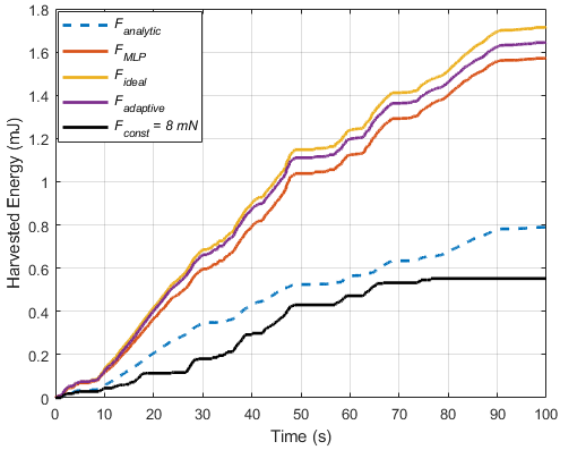


Figure 80. *Harvested energy as a function of time using the MLP mapping, analytical approach, and the proposed adaptive method.*

Figure 80 displays the harvested energy obtained by using this adaptive approach compared with the MLP, as well as a purely analytical approach, for a given scenario. Our proposed adaptive estimation approach results in a gain of approximately 11 % in the harvested energy. Also shown in Figure 80 are the harvested energy curves for a constant value of F as well as its real optimal value (denoted by F_{ideal}, obtained through exhaustive search). The result using F_{ideal} serves as an upper bound on the achievable harvested energy for a given acceleration data.

It should be noted that the numerical values of the gains reflected in Figure 80 are for a CFPG device with a frame size of 15×15×1.5 mm3. The dependency of these gains on the length of the decision interval, i.e., Δ should also be studied. This could also depend on the dynamics of the micro-harvester acceleration as the result of the human body movement. It is conjectured that another layer of adaptation for this decision interval could also contribute to further maximization of the achieved power. We realize that the addition of the proposed adaptive estimation module will require some operational energy itself. This energy should be a small fraction of the achieved again in the harvested power. Detailed studies on the net power gain considering the added complexity of the holding force estimator will be conducted in the future.

[1] M. Dadfarnia, K. Sayrafian, P. Mitcheson, and J. S. Baras. Maximizing Output Power of a CFPG Micro Energy-Harvester for Wearable Medical Sensors. In *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare* (MOBIHEALTH), Athens, Greece, November 3-5, 2014.

[2] M. Roudneshin, K. Sayrafian, and A. Aghdam. A Machine Learning Approach to the Estimation of Near-Optimal Electrostatic Force in Micro Energy-Harvesters. In *Proceedings of the 7th Annual IEEE International*

Conference on Wireless for Space and Extreme Environments (WISEE 2019), Ottawa, Canada, October 16-18, 2019.

[3] M. Roudneshin, K. Sayrafian, and A. G. Aghdam. Adaptive Estimation of Near-Optimal Electrostatic Force in Micro Energy-Harvesters. In *Proceedings of the IEEE Conference on Control Technology and Applications* (CCTA), Montreal, Canada, August 24-26, 2020.

Combinatorial Testing for Software based Systems

Raghu N. Kacker
D. Richard Kuhn (NIST ITL)
Yu Lei (University of Texas at Arlington)
Dimitris E. Simos (SBA-Research, Austria)
Mohammad S. Raunak (NIST ITL)
James F. Lawrence
Eric Wong (University of Texas at Dallas)
Itzel Dominguez-Mendoza (CENAM, Mexico).
Sebastian Barillaro (INTI, Argentina)

https://csrc.nist.gov/projects/automated-combinatorial-testing-for-software

In 1997, the Mars Pathfinder began experiencing system resets at seemingly unpredictable times soon after it landed and began collecting data. Fortunately, engineers were able to deduce and correct the problem, which occurred only when (1) a particular type of data was being collected, and (2) intermediate priority tasks exceeded a certain load, resulting in a blocking condition that eventually triggered a reset. Situations of this type are known as *interaction faults*. Many real-time failures of software systems have been traced to such faults. These are often insidious in that they may remain hidden until the unfortunate combination is encountered during system operation.

Combinatorial testing (CT) is a versatile methodology for detecting interaction faults. CT began as pairwise (2-way) testing in which all pairs of the test values for all pairs of test factors are checked. Thus, pairwise testing can detect faults involving single factors or interactions between two factors. CT is based on an empirical observation, referred to as the *interaction rule*, that while the behavior of a software system may be affected by many factors, only a few are involved in any given failure. NIST investigations of failures in actual systems have shown that while most faults involved a single factor or interaction between two factors, some faults involved three or more factors [1]. (A fault involving more than six factors has not yet been reported.) Thus, while pairwise testing is useful, it may not be adequate for detecting interaction faults involving more than two test factors.

More than a decade ago, NIST took the initiative to extend pairwise (2-way) CT to higher strength t-way CT for $t > 2$. NIST has helped make CT practical by developing research tools and techniques for generating combinatorial test suites. CT has now gained significant interest from the international software testing community. Many successful results from the use of CT in aerospace, automotive, and financial service industries, as well as defense, security, and electronic medical systems have since been reported.

A suite of test cases for combinatorial t-way testing includes (covers) at least once all possible t-tuples of the test values for every set (combination) of t factors out of the complete set of all k factors that are tested ($k > t$). Use of mathematical objects called covering arrays makes it possible to check all t-tuples of the test values with a small number of test cases. Table 1 shows a covering array of 13 rows and 10 columns each having two possible value 0 and 1. Columns correspond to the factors and the rows correspond to the test cases. The number of possible sets (combinations) of 3 out of 10 test factors is $(10 \times 9 \times 8)/(3 \times 2 \times 1) = 120$. When each factor has two possible values, each set of 3 factors can have $2^3 = 8$ possible triples of test values ((0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (0, 0, 1), (1, 1, 0), (1, 1, 1)). So, the total number of possible triples of values for all 10 factors is $120 \times 8 = 960$. A test suite based on Table 4 includes ("covers") at least once all 960 distinct triples of the test values of ten factors.

In practice, one wants a minimal covering array, that is, an array which covers all possible t-tuples of the test values for every set of t out of all k factors with the least number of rows (test cases). In practice, many factors have dependencies and constraints, and

Table 4. *A covering array of 13 rows includes all eight triplets (000, 001, 010, 011, 100, 101, 110, and 111) of the possible values (0 and 1) for every one of the 120 possible sets of 3 out of 10 test factors represented by the columns (for example, see colored entries)*

	Columns									
Rows	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	0	1	0	0	0	0	1
4	1	0	1	1	0	1	0	1	0	0
5	1	0	0	0	1	1	1	0	0	0
6	0	1	1	0	0	1	0	0	1	0
7	0	0	1	0	1	0	1	1	1	0
8	1	1	0	1	0	0	1	0	1	0
9	0	0	0	1	1	1	0	0	1	1
10	0	0	1	1	0	0	1	0	0	1
11	0	1	0	1	1	0	0	1	0	0
12	1	0	0	0	0	0	0	1	1	1
13	0	1	0	0	0	1	1	1	0	1

hence not all combinations of the test values may be logically or physically valid. A combinatorial test suite must avoid such forbidden combinations. Generating minimal covering arrays that avoid forbidden combinations is a difficult computational problem [2]. A great deal of research has been done to develop mathematical and computational methods to generate minimal covering arrays of this type. NIST and its collaborators have developed several such algorithms.

NIST-Developed Tools. NIST has developed several research tools to make CT practical. ACTS (for Automated Combinatorial Testing for Software), which was developed in cooperation with the University of Texas at Arlington, includes several algorithms to generate high strength test suites for CT. The ACTS algorithms are optimized to efficiently avoid forbidden combinations of test settings. More than 3950 users have downloaded executable versions of the ACTS algorithms from the NIST webpage for CT. (It is difficult to ascertain the number of users because some users have redistributed to others and some are students who may have used it only once for a single project.)

A second research tool, CCM (for Combinatorial Coverage Measurement), developed jointly by NIST and a guest researcher from CENAM, the national metrology institute of Mexico, describes the incompleteness of a test suite that may not have been developed from a CT viewpoint. Basic combinatorial coverage measurements describe the incompleteness of a test suite relative to a test suite based on a covering array that includes all possible *t*-tuples of values for every *t*-factor combination for various values of *t*. The combinatorial deficiency of a test suite can be remedied by additional tests. Thus, CCM can help guide the expansion of a test suite to satisfy stated combinatorial requirements [3]. The latest version of CCM supports constraints which exclude forbidden combinations of values. A parallel processing version is also available.

Impact of NIST Research. NIST efforts have sparked a surge of research and application of combinatorial testing technology. A 2010 NIST Special Publication on CT was downloaded more than 30 000 times by the end 2014 [4]. In 2013, we published a book with CRC Press on this topic [5]. One of the first large-scale users that we worked with is a group at the U.S. Air Force Base in Eglin, Florida. The behavior of one of their systems depended on the sequential order of certain events. This led to the problem of testing sequences of events, which required development of new mathematical objects called *sequence covering arrays* [6, 7, 8]. Lockheed-Martin, a large U.S. defense contractor, reported (based on eight projects) that use of CT reduced cost of testing by about 20 % with 20 % to 50 % improvement in test coverage [9]. CT methods are now being used in diverse areas such as financial services, automotive, automation, avionics, video coding standards, and for security testing. The NIST webpage for CT cites over forty application papers. For testing a software-based system, no single approach is enough. Plural approaches are generally needed at various stages of software development and installation. CT complements other approaches for testing, verification, and validation of software-based systems. CT is now included in software engineering courses taught in more than 18 U.S. universities and 2 Austrian universities at least. NIST efforts on technology transfer of CT tools and techniques received the 2009 Excellence in Tech Transfer Award from the Federal Laboratory Consortium-Mid Atlantic Region.

CT has also gained significant interest from the research community. In 2012, NIST took lead in organizing a workshop on CT[36] in conjunction with the IEEE International Conference on Software Testing, Verification, and Validation (ICST), a premier conference in this field. Since then, an International Workshop on Combinatorial Testing (IWCT) has become an annual event for sharing advancements in CT tools and techniques, as well as results from practical industrial use of CT. The ninth such IWCT[37] was held (virtually because of the worldwide COVID-19 pandemic) on October 24, 2020 in conjunction with ICST 2020.[38] Four of us (Kacker, Kuhn, Lei, and Simos) were among the co-organizers. IWCT 2020 received 12 submission, from which the Program Committee selected 8 papers (5 full, 2 short, and 1 extended abstract). The program contained research papers as well as on applications and use cases of CT.

Recent Accomplishments.

- *Fault localization* is a generic problem in testing software-based systems. It refers to detecting from the pass/fail results of testing an error (bug) that triggers a failure. Our research paper on tools and techniques for fault localization from combinatorial testing appeared in the IEEE Transactions on Software Engineering [10].
- *Practical issues in using CT*: We investigated effectiveness and challenges in CT relative to various types of functional testing, expert-judgment testing, and random testing in five real-life studies. CT is effective in detecting multi-factor faults. The biggest challenge is input space modeling [11].
- *Dataset reduction for AI testing*: We investigated the effect of using datasets reduced by random and stratified sampling to speed up testing machine learning algorithms. Reduced datasets achieve branch coverage identical or like the original datasets. Branch coverage could be used as an

[36] http://www.research.ibm.com/haifa/Workshops/ct2012/

[37] http://gist.nju.edu.cn/iwct2020/

[38] https://icst2020.info/

alternative to the more expensive mutation coverage in testing machine learning algorithms [12].

- *Input space coverage*: Structural metrics such as statement and branch coverage must be supplemented with measures of the input space coverage to provide a means of verifying that an adequate input model has been defined [13].
- *Combinatorial methods for explainable AI*: We investigated use of methods derived from combinatorial fault interaction localization to produce explanations or justifications for decisions made by artificial intelligence and machine learning (AI/ML) systems [14].
- *CAgen: a fast-combinatorial test generation tool with support of constraints*: CT is being used for hundreds or thousands of factors (parameters) successfully. Generating combinatorial test suites for such large complex systems is computationally challenging. Our collaborators (in SBA-Research, Austria) have developed a tool CAgen that is significantly faster than the alternatives and supports constraints handling and higher-index arrays [15]. Both command line and GUI versions are offered.
- *Vulnerability trends in web servers and browsers:* Five categories of vulnerabilities show a gradual downward trend with substantial variability from year to year. Vulnerability categorized as "information leak" seem to be showing a slight upward trend over the last couple of years. For web browsers, there persists the disappointing trend of a clearly rising "buffer error type vulnerabilities" over the last ten years [16].
- *Low Power Wide Area Networks (LPWAN) for communications of mobile sensor data*: A foreign guest researcher Sebastian Barillaro from the National Metrology Institute (NMI) of Argentina investigated the use of an LPWAN option (called LoRaWAN) for communication in networks of sensors used for experiments in the NIST Gaithersburg campus. LoRaWAN is a stand-alone-option that does not need WiFi from the Internet Service Providers (ISPs) [17]. Sebastian was sponsored by the NIST International and Academic Affairs Office, ACMD, and the NIST Computer Security Division.
- *State of the art methods for nuclear safety systems*: We participated in a series of meetings with the US Nuclear Regulatory Commission, and personnel from universities and government organizations in Europe and Japan. Meetings were organized by national bodies concerned with the safety of software used in nuclear reactors, under the auspices of the OECD Halden Reactor Project. The meetings produced a report [18] from the OECD organization sponsoring the initiative, and follow-on recommendations developed by the Idaho National Laboratory for NRC [19]. We contributed briefings on high assurance software verification and testing, and text and recommendations to the OECD report.

Current Projects.

- *Combinatorial Fuzzing for Smart Contracts*: A blockchain is a decentralized computer system that uses cryptographic and distributed consensus technologies to record information in a permanent and verifiable manner. A smart contract is a computer program that executes one or more transactions on a blockchain. When some conditions are met, the transactions are automatically executed, and their results are recorded on the blockchain in a way that cannot be changed. We are investigating combinatorial fuzzing (a combination of CT and fuzzing) to improve code coverage and detection of vulnerabilities in testing smart contracts. Combinatorial Fuzzing requires testing tools. We have developed a prototype research tool called MagicMirror for combinatorial fuzz testing of smart contracts. We used MagicMirror to test 2,397 real-world smart contracts. MagicMirror performs much better than another academic research tool. A research paper has been submitted for presentation at a leading IEEE conference in 2021.
- *CT for Multiple Input Models*: Existing work on CT is mainly concerned with test generation for a single input model (that is, selection of test factors and their values). We are investigating test generation for multiple input models with shared parameters (test factors). This CT-MM problem arises in finite-state-machine-based testing and graphical user interface testing.
- We built a prototype research tool and applied it to five real-life Android applications. We summarized our work in a submission to the IEEE Transactions on Software Engineering (TSE). In response to comments received from the first round of TSE reviews, we conducted additional experiments and revised the paper, and resubmitted for the second round of reviews.
- *CT for Deep Learning*: We are investigating CT for deep learning models used in autonomous vehicles. Deep learning models are used to perform tasks such as pedestrian detection, obstacle recognition, and steering control. It is important to rigorously test these models to ensure safety. Our preliminary results from experiments with three deep learning models suggest that CT could generate test images to effectively identify inconsistent behavior among the models. A research paper has been submitted to an IEEE conference in 2021.
- *Combinatorial Security Testing for IoT systems*: Out collaborators from the SBA-Research are leading an investigation to test for the security of

Internet of Things (IoT) home automation systems using combinatorial methods. We created input parameter models to be used with a combinatorial test case generation strategy. We evaluated the generated test cases on a real-world IoT appliance settings compatible with the created abstract models. We implement an automated test execution framework which includes automated test oracles for evaluation purposes. A paper has been submitted to a leading IEEE conference in 2021.

- *A CT and Graph-Based Technique for testing Internet of Things (IoT)*: Our collaborators from the University of Texas at Dallas are leading an investigation to test IoT systems using CT and graph-based techniques. We developed (1) an improved and enhanced model for testing IoT, (2) applied the developed IoT testing framework on two real-world IoT systems, (3) developed a prototype research tool to assist practitioners in applying CT-IoT testing framework to industrial IoT systems. A paper entitled "A Graph-Based Combinatorial Testing Framework for Effective IoT Testing is being finalized for submission.

[1] D. R. Kuhn, D. R. Wallace, and A. M. Gallo, Jr. Software Fault Interactions and Implications for Software Testing. *IEEE Transactions on Software Engineering* **30** (2004). 418-421. DOI: 10.1109/TSE.2004.24

[2] L. Kampel and D. E. Simos. A Survey on the State of the Art of Complexity Problems for Covering Arrays. *Theoretical Computer Science* **800** (2019),107-124. DOI: 10.1016/j.tcs.2019.10.019

[3] D. R. Kuhn, R. N. Kacker and Y. Lei. Combinatorial Coverage as an Aspect of Test Quality. *CrossTalk: The Journal of Defense Software Engineering* (March/April 2015), 19-21.

[4] D. R. Kuhn, R. N. Kacker, and Y. Lei. Practical Combinatorial Testing. NIST Special Publication 800-142, October 2010. URL: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-142.pdf

[5] D. R. Kuhn, R. N. Kacker, and Y. Lei. *Introduction to Combinatorial Testing.* CRC Press, 2013.

[6] D. R. Kuhn, J. M. Higdon, J. F. Lawrence, R. N. Kacker and Y. Lei. Combinatorial Methods for Event Sequence Testing. In *Proceedings of the 5-th IEEE International Conference on Software Testing, Verification and Validation Workshops* (ICSTW), Montreal, Canada, April 17-21, 2012, 601-609. DOI: 10.1109/ICST.2012.147

[7] D. R. Kuhn, J. M. Higdon, J. F. Lawrence, R. N. Kacker and Y. Lei. Efficient Methods for Interoperability Testing using Event Sequences. *CrossTalk: The Journal of Defense Software Engineering*, July/August 2012, 15-18. URL: https://apps.dtic.mil/docs/citations/ADA566540

[8] Y. M. Chee, C. J. Colbourn, D. Horsley, and J. Zhou. Sequence Covering Arrays. *SIAM Journal of Discrete Mathematics* **27** (2013), 1844-1861. DOI: 10.1137/120894099

[9] J. Hagar, D. R. Kuhn, R. N. Kacker, and T. Wissink. Introducing Combinatorial Testing in a Large Organization. *IEEE Computer* **48** (April 2015), 64-72. DOI: 10.1109/MC.2015.114

[10] L. S. Ghandehari, Y. Lei, R. N. Kacker, D. R. Kuhn, T. Xie, and D. Kung. A Combinatorial Testing-Based Approach to Fault Localization. *IEEE Transactions on Software Engineering* 46 (2020), 616-645. DOI: 10.1109/TSE.2018.2865935

[11] L. Hu, W. E. Wong, D. R. Kuhn, and R. N. Kacker. How Does Combinatorial Interaction Testing Perform in the Real World: An Empirical Study on Five Industrial Systems. *Journal of Empirical Software Engineering* **25** (2020), 2661-2694. DOI: 10.1007/s10664-019-09799-2

[12] J. Chandrasekaran, H. Feng, Y. Lei, R. N. Kacker, and D. R. Kuhn. Effectiveness of Dataset Reduction in Testing Machine Learning Algorithms. *In Proceedings of IEEE Artificial Intelligence Testing* (AI Test), 133-140, 2020. DOI: 10.1109/AITEST49225.2020.00027

[13] D. R. Kuhn, R. N. Kacker, Y. Lei, and D. E. Simos. Input Space Coverage Matters. *IEEE Computer* **55** (2020), 37-44. DOI: 10.1109/MC.2019.2951980

[14] D. R. Kuhn, R. N. Kacker, Y. Lei, and D. E. Simos. Combinatorial Methods for Explainable AI. In *Proceedings of the 2020 IEEE International Conference on Software Testing, Verification and Validation Workshops* (ICSTW), Online, October 24-28, 2020, 167-170. DOI: 10.1109/ICSTW50294.2020.00037

[15] M. Wagner, K. Kleine, D. E. Simos, D. R. Kuhn, and R. N. Kacker. CAGEN: A Fast Combinatorial Test Generation Tool with Support for Constraints and Higher-Index Arrays. In *Proceedings of the 2020 IEEE International Conference on Software Testing, Verification and Validation Workshops* (ICSTW-2020), Online, October 24-28, 2020, 191-200. DOI: 10.1109/ICSTW50294.2020.00041

[16] M. S. Raunak, D. R. Kuhn, R. Kogut, and R. N. Kacker. Poster: Vulnerability Trends in Web Servers and Browsers. In *Proceedings of the Symposium on the Science of Security* (HotSoS). Online, September 22-23, 2020. DOI: 10.1145/3384217.3384227

[17] S. Barillaro, S. Rhee, G. Escudero, R. N. Kacker, M. L. Badger, and D. R. Kuhn. Low-Power Wide Area Networks (LPWAN) for Communications of Mobile Sensor Data. *In Proceedings of the 2nd ACM/EIGSCC Symposium on Smart Cities and Communities* (SCC 2020), Portland, Oregon, September 10-12, 2019, 6. DOI: 10.1145/3357492.3358629

[18] OECD Halden Reactor Project. HWR-1288. In *Workshop Proceedings: State of the Art to Assure a Digital Safety System*. October 26, 2020.

[19] K. Vedros, A. Al-Rashdan, and R. England. *Protecting Against Hazards from Common Causes in Digital Instrumentation and Control Systems.* INL/LTD-20-60088, Idaho National Laboratory, October 7, 2020.

Mathematical Knowledge Management

We work with researchers in academia and industry to develop technologies, tools, and standards for representation, exchange, and use of mathematical data. Of particular concern are semantic-based representations which can provide the basis for interoperability of mathematical information processing systems. We apply these representations to the development and dissemination of reference data for applied mathematics. The centerpiece of this effort is the Digital Library of Mathematical Functions, a freely available interactive and richly linked online resource, providing essential information on the properties of the special functions of applied mathematics, the foundation of mathematical modeling in all of science and engineering.

Digital Library of Mathematical Functions

Barry I. Schneider
Daniel W. Lozier
Bruce R. Miller
Bonita V. Saunders
Howard S. Cohl
Marjorie A. McClain
Ronald F. Boisvert
Adri B. Olde Daalhuis (University of Edinburgh)
Charles W. Clark (NIST PML)
Brian Antonishek (NIST EL)

http://dlmf.nist.gov/

Progress in science has often been catalyzed by advances in mathematics. More recently, developments in the physical sciences, such as investigations into string theory, have influenced pure mathematics. This symbiotic relationship has been extremely beneficial to both fields. Mathematical developments have found numerous applications in practical problem-solving in all fields of science and engineering, while cutting-edge science has been a major driver of mathematical research**.** Often the mathematical objects at the intersection of mathematics and physical science are mathematical functions. Effective use of these tools requires ready access to their many properties, a need that was capably satisfied for more than 50 years by the *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, which was published by the National Bureau of Standards (NBS) in 1964 [1].

The 21st century successor to the NBS Handbook, the freely accessible online Digital Library of Mathematical Functions (DLMF) together with the accompanying book, the *NIST Handbook of Mathematical Functions* [2], published by Cambridge University Press in 2010, are collectively referred to as the DLMF. The DLMF continues to serve as the gold standard reference for the properties of what are termed the special functions of applied mathematics.

The DLMF has considerably extended the scope of the original handbook as well as improving accessibility to the worldwide community of scientists and mathematicians. To cite a few examples, the new handbook contains more than twice as many formulas as the old one, coverage of more functions, in more detail, and an up-to-date list of references. The website covers everything in the handbook and much more: additional formulas and graphics, math-aware search, interactive zooming and rotation of 3D graphs, internal links to symbol definitions and cross-references, and external links to online references and sources of software.

In order to help assess the impact of the DLMF project, the NIST library has undertaken a project to track citations to the DLMF, as well as to the original NBS Handbook. While the original Handbook still receives an enormous number of citations, citations to the DLMF are steadily growing in relation to the original handbook. Google Scholar now reports more than 6,258 citations to the DLMF, a roughly 6% increase from 2019. The number of website pages served up also increased by 5% but the total number of visitors to the DLMF was essentially unchanged from 2019.

Today's DLMF is the product of many years of effort by more than 50 contributors. Its initial release in 2010, however, was not the end of the project. Corrections to errors, clarifications, bibliographic updates, and addition of new material all need to be made on a continuing basis. And, new chapters covering emerging subject areas need to be added, to assure the continued vitality of the DLMF deep into the 21st century. Since December of 2019, there were five DLMF releases, 1.0.25 (2019-11-15), 1.0.26 (2020-03-15), 1.0.27 (2020-06-15), 1.0.28 (2020-09-15), and 1.1.0 (2020-12-15) which kept us on our quarterly release schedule. The last release, 1.1.0, was a major revision as evidenced by the version number now at 1.1.

The updating of various DLMF chapters and the development of new ones continues. These include a new chapter on Several Variable Orthogonal Polynomials (SVOP) and substantial updates to the chapters on Orthogonal Polynomials (OP), Algebraic Methods (AM), and Painlevé Transcendents (PT). Four authors were identified to carry out the work. Drafts are now available for three of the chapters and are being internally reviewed. External validation of the chapters will follow in much the same manner as the original DLMF. It is

expected that the revisions of the AM and OP chapters, which are interdependent, will be released in 2021.

One of the design goals for the DLMF was that each formula would be connected to a proof in the literature. This data, visible as annotations on the website, provides either a proof for the formula, a reference to the proof for the formula or, for definitions, a reference which gives that definition. Unfortunately, this information was not provided in all cases. We are working to systematically verify the completeness and traceability to published proofs for DLMF formulae. This audit has been completed for Chapter 25 (Zeta and Related Functions) and is actively continuing for Chapters 1-5 and 22-30. In addition, work has progressed to make such information visible in metadata at the equation level. Improvements in markup to more accurately identify semantics have been developed and deployed (see Chapter 9, Airy and Related Functions).

Other recent advances include the following.

- After migrating from CVS to github, we continue to refine our practices in using the repository, especially the github's capabilities to track errors and corrections as well as desired enhancements.
- Changes to the DLMF are made via targeted github pull-requests, each of which is reviewed by other project members before inclusion.
- To facilitate the use of DLMF as a reference, we have had a long-standing commitment that every formula, table, section, etc. have a unique and permanent reference number and associated URL. Desired additions and the pending inclusion of new and revised chapter additions required a new strategy for inserting reference numbers between two existing ones, without revising their associated numbers or URLs. To do this, a new numbering scheme has been developed which treats the reference numbers as decimals rather than as integers, using a "_" as a decimal separator. See (4.21.1_5) which now separates (4.21.1) and (4.21.2) for an example. This major step resulted in labelling the last DLMF release as 1.1.0

[1] M. Abramowitz and I. Stegun, eds. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Applied Mathematics Series 55, National Bureau of Standards, Washington, DC 1964.

[2] F. Olver, D. Lozier, R. Boisvert and C. Clark, eds. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

Visualization of Complex Functions Data

Bonita Saunders
Bruce Miller
Brian Antonishek (NIST EL)
Sandy Ressler
Qiming Wang (NIST retired)

Definitions, recurrence relations, differential equations, integrals, and asymptotic expansions can provide crucial information for understanding complex mathematical functions arising in application areas of the mathematical and physical sciences, but accurate interactive visualizations can sometimes provide additional clarity as demonstrated in Figure 81.

Since the DLMF provided the impetus for this project, our work was initially driven by a quick results-oriented approach. Now that the DLMF is well established we are re-examining some graphs and visualizations to determine how they might be enhanced. For example, currently all DLMF visualizations represent complex or real valued function surfaces. However, there are some 3D static images, such as a figure showing the branch point of eigenvalues of Mathieu's equation,[39] or the bifurcation sets and sheets of surfaces in Chapter 36, Integrals with Coalescing Saddles, whose structure and meaning might be better illustrated by an interactive display.

We also continue to look for opportunities to generalize our work to benefit the larger research community. Our current work on adaptive meshes to improve our underlying computational grids supports our design process, but it may also interest other researchers in the fields of mesh generation, optimization, approximation theory, or any area related to the design of curves and surfaces for mathematical or physical applications. For this reason, we continue to publish and present our results at appropriate venues [1-5].

As time permits, we will explore opportunities to increase usability. Some DLMF chapters display chapter related thumbnail images on the title page. Creating a gallery of images for all chapters would add visual interest to the DLMF. Each image would link to a short descriptive sidebar that could include links to the chapter's application section or related function visualizations. New and updated chapters currently under development can provide a test bed for this work.

[1] B. Saunders. Complex Variables, Mesh Generation, and 3D Web Graphics: Research and Technology Behind the Visualizations in the NIST Digital Library of Mathematical Functions. In *Proceedings of the Golden Anniversary Celebration of the National Association of Mathematicians,* AMS Contemporary Mathematics Series **759** (2020). DOI: 10.1090/conm/759/15272 .

[39] https://dlmf.nist.gov/28.7#F1

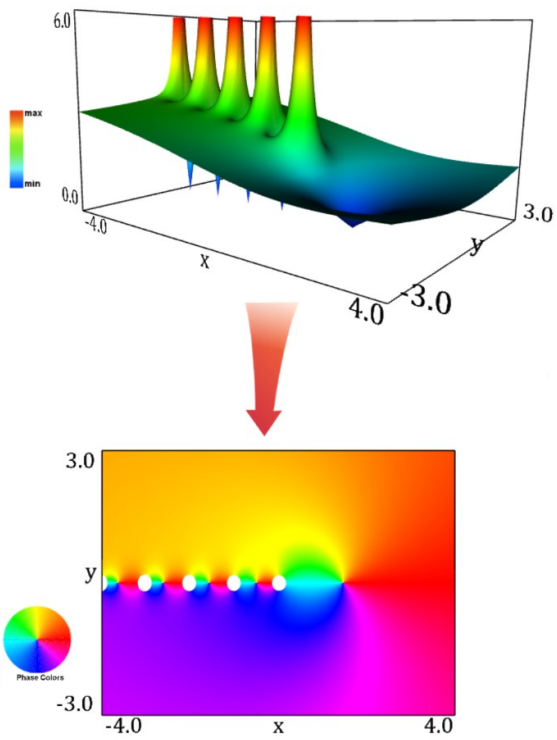


Figure 81. *DLMF visualizations provide several options for exploring mathematical function surfaces. The complex psi, or digamma, function surface, $|\psi(x+iy)|$, shown at the top with a height based color map, can be transformed into a density plot by viewing the surface from above and scaling it down to 0 in the vertical direction. Going a step further, changing the color map to a phase, or argument, based map yields the phase density plot below it. The white circles along the real axis (y=0) indicate the location of poles. The poles are interspersed with points representing zeros of the function. Note that the progression of phase colors around the poles is the reverse of colors around the zeros, indicating a phase change of opposite sign. This is analogous to fluid flow generated by a semi-infinite line of vortices with alternating sign of circulation, resulting in the near cancellation of flow seen far from the real axis.*

[2] B. Saunders, "From Abramowitz and Stegun to the NIST Digital Library of Mathematical Functions and Beyond." History of Mathematics Class, Prof. Hortensia Soto, Colorado State University, Online, November 10, 2020.

[3] B. Saunders. "Mathematics, Mesh Generation, & 3D Graphics on the Web and Finding a Career at NIST." MD-DC-VA Fall Section Meeting, Mathematical Association of America, Norfolk State University, November 9, 2019.

[4] B. Saunders. "B-Splines and 3D Web Graphics: Mathematics and Technology Behind the Visualizations of the NIST Digital Library of Mathematical Functions." George Mason University, Fairfax, VA, October 25, 2019.

[5] B. Saunders. "B-Spline Mesh Generation, 3D Web Graphics and the NIST Digital Library of Mathematical Functions." Lafayette College, Easton, PA, September 25, 2019.

Towards a Machine-Readable Digital Library of Mathematical Functions

Bruce Miller
Deyan Ginev (Chakra Consulting)
Tom Wiesing (University of Erlangen, Germany)

The principal initial goal of the Digital Library of Mathematical Functions (DLMF) project was to present useful definitions, relationships, properties and other information about the special functions of applied mathematics for use by scientists and engineers; that is, directly to people. In the medium term, we have striven to develop procedures for DLMF enhancements, maintenance and corrections. While we feel we have been successful so far, the longer-term goal of putting all data into a fully machine-readable form has been slower in coming, given the scale and complexity of the material. This form would improve the findability, accessibility, reuse, and, in particular, computability of the mathematics.

While one can imagine a sufficiently rich set of semantic LaTeX macros allowing the author to specify not only the printed form of the formula, but completely specify the exact semantic meaning of every symbol and expression as well, this approach is so cumbersome as to be impractical. Conversely, one might imagine eventual artificial intelligence able to infer the meaning from any given TeX markup along with the document's full text and context, but that is hardly feasible as yet. We seek the sweet spot between those extremes where the author supplies just enough clarity via semantic macros so that follow-up analysis can resolve any remaining ambiguities.

Fully implementing this strategy is still very much a work in progress, particularly the latter inference aspect, even scaled down. Progress is being made in several fundamental components, however, and there are many benefits to even partly annotated and disambiguated data in terms of web services and information discovery.

Short of fully explicit author markup, the mathematical formula must be parsed at least to determine the expression structure. We are cataloging the range of mathematical notations both within and outside of DLMF and have continued to enhance the mathematical grammar to increase the coverage. Additionally, we are exploring parsing frameworks which provide multiple parses to cover cases where the notation has several possible interpretations. Techniques for pruning these possibilities according to consistency or partial knowledge are being developed.

We have developed and made use internally a set of semantic-preserving macros for DLMF's special functions and the more notationally complex concepts.

Focusing here resolves some of the most glaring ambiguities, both syntactic and semantic. (There are so many different f's!) This even provides a convenience to authors to simplify and standardize the typing required. As the benefits of machine-readable mathematical data on the web are realized, a much broader audience of authors is likely to find these tools useful. For our macro set to be more palatable to that audience, it needs to be appropriately simplified and yet generalized. We are working towards this with the goal of publishing and making available our semantic macro set.

A common basis is also necessary for effective interoperability. Are the functions treated in DLMF actually the same as those in Mathematica, or Maple or the NAG libraries? As it turns out, usually they are, but sometimes not, and the differences can be subtle. Establishing these correspondences is exactly the purpose of the Special Function Concordance activity of the International Mathematical Knowledge Trust (IMKT) established by the Global Digital Mathematical Library (GDML) working group of the International Mathematical Union. We are participating in this effort to establish this concordance between the various special functions, in all their flavors, as used within various handbooks, such as the DLMF, and software systems, such as Mathematica, Maple and NAG, to assure interoperability.

For our part, we are developing a catalog of the special functions as defined in DLMF. We start by listing formulas which we consider to be the defining ones for each function. However, the key is to focus on those aspects, those choices, that can distinguish our version of a particular function from those of other systems. Choices of argument conventions (e.g., elliptic functions $\mathrm{sn}(u,k)$ vs. $\mathrm{sn}(u,m)$) are significant. Patterns of singularities and type signatures help distinguish different extensions and generalizations of functions. A trickier class of differences are choices made for the location of branch-cuts or which value is taken on the cuts, or indeed the choice made to avoid them entirely as multivalued functions. Given this set of distinguishing features, we are collecting and encoding that information for each of the DLMF's functions in the form of OpenMath Content Dictionaries.

DLMF Standard Reference Tables on Demand

Bonita Saunders
Bruce Miller
Marjorie McClain
Annie Cuyt (University of Antwerp)
Stefan Becuwe (University of Antwerp)
Franky Backeljauw (University of Antwerp)
Sean Brooks (Coppin State University)
Ron Buckmire (Occidental College)
Rachel Vincent-Finley (Southern U. and A&M College)
Christopher Schanzle

http://dlmftables.uantwerpen.be/

The advent of reliable computing machines, computer algebra systems, and multiple precision computational packages diminished the need for tables of reference values for computing function values by interpolation, but today's numerical analysts, scientific researchers, and software developers still need a way to test various software implementations for computing mathematical function values. DLMF Standard Reference Tables on Demand (DLMF Tables) is a collaborative project between ACMD and the University of Antwerp Computational Mathematics Research Group (CMA) [1-5] designed to address this problem. The goal is to develop an online system where users can generate tables of special function values at user-specified precision with an error certification to test their own algorithms or confirm the accuracy of results from a commercial or publicly available package.

The DLMF Tables team has already developed a beta site at the University of Antwerp, based on CMA's MpIeee, a multiprecision IEEE 754/854 compliant C++ floating point arithmetic library, but a permanent site at NIST is planned. A needed boost to the project occurred this past summer when B. Saunders was asked to lead a project of her choice at the African Diaspora Joint Mathematics Workshop[40] (ADJOINT 2020) sponsored by the Mathematical Sciences Research Institute (MSRI) in Berkeley, California [6]. The aim of the two-week virtual workshop was to encourage collaborative work in the mathematical sciences between established PhD researchers and less experienced researchers, especially U.S. citizens from the African Diaspora, that is, descendants of people from Sub-Saharan African countries. Saunders chose a project that introduced her three-member ADJOINT group to Validated Numerical Computations of Mathematical Functions, a field that includes the error analysis and code development needed to compute certifiably accurate function values. Since MSRI's financial support of the group covers an entire year, post-workshop contact has continued with

[40] https://www.msri.org/web/msri/scientific/adjoint

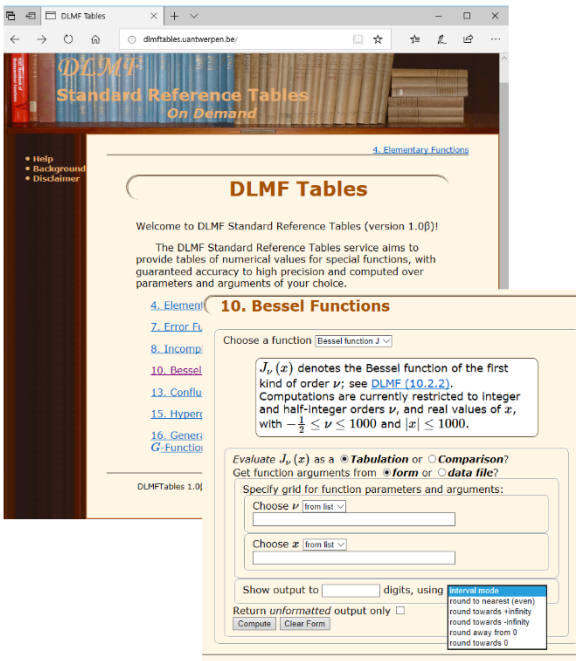


Figure 82. *DLMF Tables generates tables of special function values at user specified precision. Users input real values and parameters where the function is to be evaluated. The user may request output in interval mode, where the output is shown as a table of intervals that bound the true results or may request output in one of several rounding modes. Users may also choose to compare their own table of values to the reference values generated by the system.*

biweekly videoconference meetings to discuss and verify results in key references, including Higham [8] and CMA [1-5]. The ADJOINT group's hard work led to plans for three talks at the 2021 Joint Mathematics Meetings. The group is also working on a paper that will be co-authored by members of the ACMD and U. Antwerp groups [7].

Until now, ACMD has primarily been responsible for the front-end interface for DLMF Tables, while CMA has been responsible for the back-end computational engine built around the error analysis and MpIeee library. The work with the ADJOINT group is a first step toward bringing these capabilities together so that CMA's function computation codes can be maintained and eventually, further developed here at NIST.

[1] F. Backeljauw, S. Becuwe, A. Cuyt, J. Van Deun, and D. Lozier. Validated Evaluation of Special Mathematical Functions. *Science of Computer Programming* **10** (2013), 1016.

[2] M. Colman, A. Cuyt and J. Van Deun, Validated Computation of Certain Hypergeometric Functions. *ACM Transactions on Mathematical Software* **38:2** (January 2012), Article 11.

[3] F. Backeljauw. A Library for Radix-independent Multiprecision IEEE-compliant Floating-point Arithmetic. Technical Report 2009-01, Department of Mathematics and Computer Science, Universiteit Antwerpen, 2009.

[4] A. Cuyt, V. B. Petersen, B. Verdonk, H. Waadeland and W. B. Jones. *Handbook of Continued Fractions for Special Functions*. Springer, New York, 2008.

[5] A. Cuyt, B. Verdonk, H. Waadeland. Efficient and Reliable Multiprecision Implementation of Elementary and Special Functions. *SIAM Journal of Scientific Computing* **28** (2006), 1437-1462.

[6] B. Saunders, S. Brooks, R. Buckmire, R. Vincent-Finley. "Validated Numerical Computation of Mathematical Functions." African Diaspora Joint Mathematics Workshop (ADJOINT), Online, June 26, 2020.

[7] B. Saunders, S. Brooks, R. Vincent-Finley, R. Buckmire, F. Backeljauw, S. Becuwe, B. Miller, M. McClain, and A. Cuyt. Validated Computation of Special Mathematical Functions. In preparation.

[8] N. Higham. *Accuracy and Stability of Numerical Algorithms*. Second edition. Society for Industrial and Applied Mathematics, Philadelphia, 2002.

NIST Digital Repository of Mathematical Formulae

Howard S. Cohl
Marjorie A. McClain
Bonita V. Saunders
Abdou Youssef
Moritz Schubotz (University of Konstanz)
André Greiner-Petter (Technische Universität Berlin)
Alan Sexton (University of Birmingham)
Avi Trost (Brown University)
Rajen Dey (University of California Berkeley)

The NIST Digital Repository of Mathematical Formulae (DRMF) is an online compendium of formulae for orthogonal polynomials and special functions (OPSF) designed to a) facilitate interaction among a community of mathematicians and scientists interested in OPSF; b) be expandable, allowing the input of new formulae from the literature; c) provide information for related linked open data projects; d) represent the context-free full semantic information concerning individual formulas; e) have a user friendly, consistent, and hyperlinkable viewpoint and authoring perspective; f) contain easily searchable mathematics; and g) take advantage of modern MathML tools for easy-to-read, scalably rendered content-driven mathematics.

Our DRMF implementation, previously built using MediaWiki (the wiki software used by Wikipedia), is currently in migration to a different software platform, namely the platform used by the NIST Digital Library of Mathematical Functions (DLMF). See Figure 83 for the current draft of the DRMF home page, and Figure 84 for a sample DRMF formula page. The DRMF has been summarized in a series of papers [1-3]. A key asset in the development of DRMF context free semantic content is the utilization of a set of LaTeX macros and

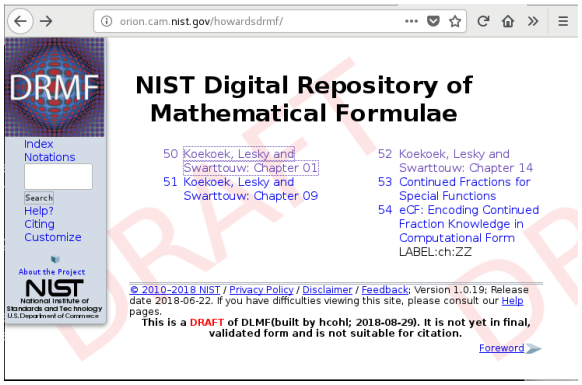


Figure 83. *Current draft of the DRMF home page displaying the current table of contents.*

macro call functionality created by Bruce Miller of ACMD to achieve the encapsulation of semantic information within the DLMF [4]. These macros give us the capability to tie LaTeX commands in a mostly unambiguous way to mathematical functions defined in an OPSF context. There are currently 540 DLMF LaTeX macros, as well as an additional 156 which have been created specifically for the DRMF. Most if not all DLMF macros have at least one DLMF web page associated with them. One goal is to have definition pages for all additional DRMF macros. The use of DLMF and DRMF macros guarantees mathematical and structural consistency in the DRMF. We refer to LaTeX source with these incorporated macros as semantic LaTeX.

DRMF formula seeding is currently focused on

1) Koekoek, Lesky, and Swarttouw (KLS) chapters 1 (Definitions and Miscellaneous Formulas), 9 (Hypergeometric Orthogonal Polynomials), and 14 (Basic Hypergeometric Orthogonal Polynomials [5];
2) Koornwinder KLS addendum LaTeX data [5];
3) Wolfram Computational Knowledge of Continued Fractions Project (eCF) [3];
4) Continued Fractions for Special Function (CFSF) Maple dataset hosted by the University of Antwerp [3,7];
5) Bateman Manuscript Project (BMP) books [8]; and
6) Magnus, Oberhettinger, and Soni (MOS) books [3,9].

For these seed projects, we are developing Python and Java software to incorporate DLMF and DRMF macros into the corresponding LaTeX source. Our coding efforts have also focused on extracting data from LaTeX source as well as generating DRMF semantic LaTeX. We have developed Java software for the seeding of the eCF and CFSF projects which involve conversion from Mathematica and Maple format to DLMF and DRMF macro incorporated semantic LaTeX [3].

In August 2014, the DRMF Project obtained permission and license to use BMP material as seed content for the DRMF from Adam Cochran, Associate General Counsel of Caltech. Caltech has loaned us copies of the BMP. In February 2018, we received permission and license to use the KLS and MOS material as seed content for the DRMF from Springer Nature. We have forwarded the copies of BMP and MOS to A. Sexton, Scientific Document Analysis Group, School of Computer Science, University of Birmingham, UK. Sexton has scanned the BMP and MOS and is developing software to perform mathematical optical character recognition to obtain LaTeX source. To enhance the development process of the OCR software, we have developed an automated testing framework that uses visual diffs of the original scanned source and the rendering of the generated LaTeX source.

Current and future DRMF MediaWiki development projects include the production of formula output representations (such as semantic LaTeX, MathML, Mathematica, Maple, and Sage); incorporation of sophisticated DLMF and DRMF macro related formula search; and the development of capabilities for user community formula input. In this vein, A. Youssef has written a grammar-based mathematical language processor (MLP) that uses JavaCC to parse mathematical LaTeX expressions [10]. Based on the MLP, A. Greiner-Petter has developed a Java tool referred to as LaCASt to convert mathematical LaTeX expressions, which contain DLMF and DRMF macros, to a given computer algebra system source format. This Java tool provides further information of the conversion about possible ambiguities and differences in definitions, domains and branch cuts between the semantic LaTeX source and the CAS source. Furthermore, it is designed to be easily extendable to other computer algebra systems and currently supports Maple and Mathematica input sources.

The following are some recent accomplishments.

- ACMD summer student A. Trost worked on the project “Semantic Enhancement for Wronskians and Prime Derivative Notation,” and ACMD summer student R. Dey worked on the project “Semantic Enhancement for Sums and Products.”
- Cohl is a member of the program committee for the 14th Conference on Intelligent Computer Mathematics, Timisoara, Romania, July 2021.
- In [11] we present a distributional analysis of mathematical formulae on arXiv and zbMATH as well as retrieve relevant mathematical objects for given textual search queries.
- In [12], which focuses on LaCASt, we present a first comprehensive approach to verify a digital mathematical and two computer algebra systems with one another by converting mathematical expressions from one system to the other. This is accomplished by our development of LaCASt which translates formulae from the DLMF to the computer algebra

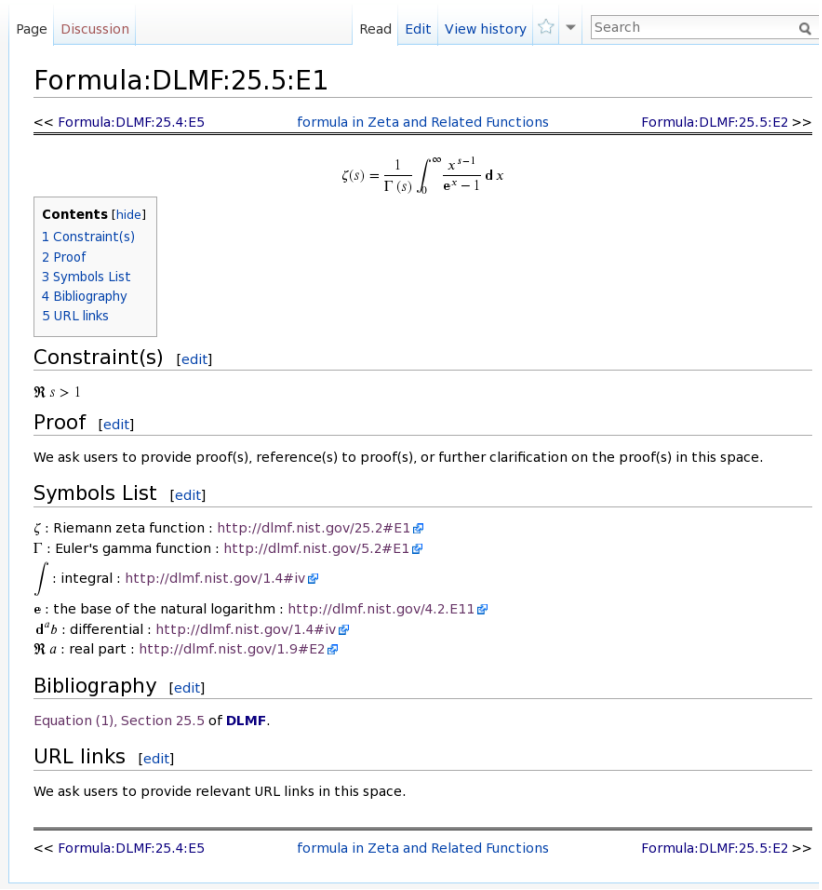


Figure 84. *Sample DRMF formula page, taken from the KLS Chapter 1 dataset.*

systems Maple and Mathematica. This tool will be actively used in the DRMF.

- The KLS datasets have been uploaded to our DLMF platform as well as the CFSF and eCF datasets.

By working with Andrea Fisher-Scherer, Rights Administrator, Artists Rights Society, New York, NY, we have received permission from Foundation Vasarely, to use an image of one of Victor Vasarely's paintings as the DRMF logo; see Figure 85.

[1] H. S. Cohl, M. A. McClain, B. V. Saunders, M. Schubotz, and J. C. Williams. Digital Repository of Mathematical Formulae. *Lecture Notes in Artificial Intelligence* **8543** (2014), Proceedings of the Conferences on Intelligent Computer Mathematics 2014, Coimbra, Portugal, July 7-11, 2014, (S. M. Watt, J. H. Davenport, A. P. Sexton, P. Sojka, and J. Urban, eds.), Springer, 419-422.

[2] H. S. Cohl, M. Schubotz, M. A. McClain, B. V. Saunders, Cherry Y. Zou, Azeem S. Mohammed, and Alex A. Danoff. Growing the Digital Repository of Mathematical Formulae with Generic LaTeX Sources. *Lecture Notes in Artificial Intelligence* **9150** (2015), Proceedings of the Conference on Intelligent Computer Mathematics 2015, Washington DC, USA, July 13-17, 2015, (M. Kerber, J. Carette, C. Kaliszyk, F. Rabe, and V. Sorge, eds.), Springer, 280-287.

[3] H. S. Cohl, M. Schubotz, A. Youssef, A. Greiner-Petter, J. Gerhard, B. V. Saunders, M. A. McClain, J. Bang, and K. Chen. Semantic Preserving Bijective Mappings of Mathematical Formulae between Word Processors and Computer Algebra Systems. *Lecture Notes in Computer Science* **10383** (2017), Proceedings of the Conference on Intelligent Computer Mathematics 2017, Edinburgh, Scotland, U.K., July 17-21, 2017, (H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke, eds.), Springer, 115-131.

[4] B. Miller. "Drafting DLMF Content Dictionaries." Open-Math Workshop, 9th Conference on Intelligent Computer Mathematics, CICM 2016, Bialystok, Poland.

[5] R. Koekoek, P. A. Lesky, and R. F. Swarttouw. *Hypergeometric Orthogonal Polynomials and their q-Analogues*. Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2010.

[6] T. H. Koornwinder. Additions to the Formula Lists in Hypergeometric Orthogonal Polynomials and their q-analogues by Koekoek, Lesky and Swarttouw. arXiv:1401.0815, June 2015.

[7] A. Cuyt, V. Petersen, H. Waadeland, W. B. Jones, F. Backeljauw, C. Bonan-Hamada, and S. Becuwe. *Handbook of Continued Fractions for Special Functions*. Springer, New York, 2008.

[8] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi. *Higher Transcendental Functions*. Vols. I, II, III, Robert E. Krieger Publishing Co., Melbourne, FL, 1981.

[9] H. S. Cohl, A. Greiner-Petter, and M. Schubotz. Automated Symbolic and Numerical Testing of DLMF Formulae using Computer Algebra Systems. *Lecture Notes in Computer Science* **11006** (2018), Proceedings of the Conference on Intelligent Computer Mathematics 2018, Hagenberg, Austria, August 13-17, 2018, (F. Rabe, W. Farmer, G.O. Passmore, A. Youssef, eds.), Springer, 39-52.

[10] A. Youssef. Part-of-Math Tagging and Applications. *Lecture Notes in Computer Science* **10383** (2017), Proceedings of the Conference on Intelligent Computer Mathematics 2017, Edinburgh, Scotland, U.K., July 17-21, 2017, (H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke, eds.), Springer, 356-374.

[11] A. Greiner-Petter, M. Schubotz, F. Müller, C. Breitinger, H. S. Cohl, A. Aizawa, and B. Gipp, Discovering Mathematical Objects of Interest—A Study of Mathematical Notations. in *WWW'20: Proceedings of the Web Conference 2020* (Y. Huang, I. King, T.-Y. Liu, and M. van Steen, eds.), April 2020, 1445-1456.

[12] A. Greiner-Petter, M. Schubotz, H. S. Cohl, M. Schubotz, A. Aizawa, B. Gipp, A. Trost, and R. Dey. Comparative

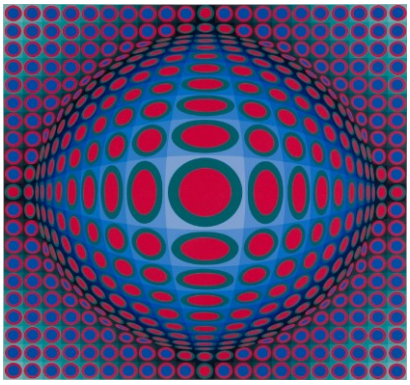


Figure 85. *DRMF logo.*

Verification of the Digital Library of Mathematical Functions and Computer Algebra Systems. In review.

Scientific Document Corpora for Natural and Mathematical Language Research

Bruce Miller
Deyan Ginev (Chakra Consulting)
Tom Wiesing (University of Erlangen, Germany)

Recent advances in machine learning have opened up new possibilities for harvesting large collections of scientific documents to find, understand and reuse information. There is much research and development still to be carried out in natural language processing (NLP) as well as extending this research to the unique characteristics of mathematical and scientific language and notation. To enable such research, it is important to have a set of documents available in a form appropriate for processing to use as both training and test data. In order to provide a statistical foundation for further work, the larger the data set the better. To that end, we have been applying our LaTeXML tool to the massive corpus at arXiv.org.[41] To demonstrate the utility of this dataset, we have carried out initial experiments on statement classification using that data set.

LaTeXML[42] was originally developed for use in converting the LaTeX sources of the Digital Library of Mathematical Functions (DLMF)[43] into web format, namely HTML and MathML. Most of arXiv.org is also in LaTeX format, albeit using a significantly less disciplined markup style and a wide variety of support packages, including uncommon or very complex ones. We have continued to develop LaTeXML, making it more robust and more accurately simulating the TeX engine. The internal infrastructure for mathematical semantics has also been enhanced, allowing us to process material using the rich siunitx and physics packages, preserving both the displayed formula as well as its mathematical intent. Moreover, we have refined the encoding of mathematics into the textual "lexemes" needed by machine learning methods to more accurately reflect each symbol's role in the text. Given the increasing variety of markup used on arXiv, as well as its fast growth, effort is required just to stay even. But with our continuing effort, we have managed to convert 96 % of arXiv's 1.6M documents into HTML+MathML with a 70 % success rate (i.e., documents producing, at worst, warnings).

Although the markup used in LaTeX documents found in the wild seldom emphasizes semantics, there are nevertheless a number of macros and environments which can be taken to reflect the author's intent. Such markup may indicate theorems, proofs, definitions as well as introductions and acknowledgments and so on. We were able to extract some 10M such annotated paragraphs in 50 categories from our conversion of arXiv [1]. Using 80 % of the documents as a training set, we discovered that many categories were too similar. For example, theorems, lemmas and propositions share language patterns. Combining these "confusion nests" yielded 13 clear cut categories into which the paragraphs could be reliably categorized with a 0.91 F1 score.

[1] D. Ginev and B. Miller. Scientific Statement Classification over arXiv.org. In *Proceedings of the 12th Language Resources and Evaluation Conference*, Online, May 2020, 1219-1226. URL: https://www.aclweb.org/anthology/2020.lrec-1.153/

Deep Learning for Math and Science Knowledge Processing

Abdou Youssef
Bruce Miller

In this project, now in its fourth year, we are developing methods and software for extracting and disambiguating semantics from text and equations in math and science documents, and applying our techniques to create new tools and applications using the latest advances in machine learning (ML), deep learning, and natural language processing. In 2020, we made considerable progress in three areas, which we describe below.

Development of a More Complete Math Dataset. We turned the contents of NIST's Digital Library of Mathematical Functions (DLMF) into a well-structured, one-of-a-kind, labeled (i.e., annotated) dataset for training and testing ML models for a number of valuable math-

[41] https://arxiv.org/
[42] https://dlmf.nist.gov/LaTeXML/
[43] https://dlmf.nist.gov/

language processing applications, and made that dataset available in the public domain. Labeled datasets are a staple of machine learning development, but are rare in math processing applications of ML, which makes our dataset unique, timely and valuable.

Math-specific Summarization. Using the above dataset, we applied the well-known page-rank and text-rank based techniques to rank the importance of individual sentences and equations, and then extract (i.e., select) from input math documents a subset of its sentences and equations as a summary of the document. That is, we can do extractive summarization of math documents. We also started exploring abstractive (i.e., "paraphrasive and fusive") summarization of math content and compared that with extractive summarization of math. Our early findings indicate that extractive methods are better than the equation portions of summaries, while abstractive methods hold a stronger promise for the pure-text portions.

Declarative Language Enhancements for Math. The source conversion tool, LaTeXML, has been continuously enhanced with new infrastructure for representing and using more complex mathematical semantics. Improvements have been made to allow more powerful declarative information (as would be needed in DLMF) as well as means of preserving what information is implicit (as would be appropriate for arXiv).

Supported by three successive annual grants from the ITL Building the Future Program, and building on the work we have done, the project will next focus on the following.

- Augment the capabilities and precision of our part-of-math tagger, so it can identify and label more types of mathematical objects in a more detailed and accurate fashion.
- Further advance the automated abstractive summarization techniques for math/STEM contents and optimize the fusion between extractive and abstractive summarization of math documents.
- Use our new dataset and the latest advances in deep learning and natural language processing for automated detection of in-text definitions of symbols and constructs and of in-text mathematical conditions and constraints on equations.
- Continue to expand our dataset to make more widely usable.
- Continue development of enhanced techniques and declarative information within the DLMF source files, leading to a fully disambiguated dataset, and to Content-MathML representation of the math that can be used in further computation.

[1] A. Youssef and B. Miller. A Contextual and Labeled Math-Dataset Derived from NIST's DLMF. In *Intelligent Computer Mathematics*, Proceedings of the 13th International Conference, CICM 2020, Online, July 26–31, 2020 (C. Benzmüller and B. Miller, eds., *Lecture Notes in Artificial Intelligence* **12236**, Springer, 1219-1226, 2020. DOI: 10.1007/978-3-030-53518-6_25

[2] A. Greiner-Petter, A. Youssef, T. Ruas, B. R. Miller, M. Schubotz, A. Aizawa, and B. Gipp. Math-word Embedding in Math Search and Semantic Extraction. *Scientometrics* **125** (2020), 3017-3046. DOI: 10.1007/s11192-020-03502-9

[3] D. Ginev and B. Miller. Scientific Statement Classification over arXiv.org. In *Proceedings of the 12th Language Resources and Evaluation Conference*, Online, May 2020, 1219-1226. URL: https://www.aclweb.org/anthology/2020.lrec-1.153/

Automated Presentation-to-Computation (P2C) Conversion

Abdou Youssef
Howard S. Cohl
Moritz Schubotz (FIZ Karlsruhe, Germany)
André Greiner-Petter (Technische U. Berlin, Germany)

One of the most obvious applications of automation is to be able to hand over to a computer a mathematical equation and expect computations to be performed with it. Yet, as simple as this sounds, it presents many challenges that only recently could be tackled. Based on the work we have been doing in math language and knowledge processing, we have developed techniques and algorithms to automate what we call the presentation-to-computation (P2C) conversion.

In collaboration with a number of internal and external researchers (including former NIST guest researchers), we were able this year to greatly enhance a system we developed earlier to translate equations from the NIST Digital Library of Mathematical Functions to the computer algebra systems Maple and Mathematica. Due to those enhancements, the accuracy of the translations was significantly improved. In future development of this translator, we will utilize deep learning techniques, making use of the advances we obtained in the aforementioned project, and utilizing the dataset we created to train deep-learning-based translators.

[1] André Greiner-Petter, Howard S. Cohl, Abdou Youssef, Moritz Schubotz, Avi Trost, Rajen Dey, Akiko Aizawa, and Bela Gipp, Comparative Verification of the Digital Library of Mathematical Functions and Computer Algebra Systems. In review.

Fundamental Solutions and Formulas for Special Functions and Orthogonal Polynomials

Howard S. Cohl
Roberto S. Costas-Santos (University of Alcala)
Hans Volkmer (University of Wisconsin-Milwaukee)
Gestur Olafsson (Louisiana State University)
Mourad E. H. Ismail (University of Central Florida)
Tom H. Koornwinder (University of Amsterdam)
James Lawrence
Jessica E. Hirtenstein (University of California Davis)
Philbert R. Hwang (University of Maryland)
Justin Park (Massachusetts Institute of Technology)
Jason Zhao (University of California Los Angeles)
Linus Ge (University of Rochester)

The concept of a function expresses the idea that one quantity (the input) completely determines another quantity (the output). Our research concerns special functions and orthogonal polynomials. A special function is a function that has appeared in the mathematical sciences so often that it has been given a name. Green's functions (named after the British mathematician George Green, who first developed the concept in the 1830s) describe the influence of linear natural phenomena such as electromagnetism, gravity, heat and waves. For example, in electrostatics, a Green's function describes the influence of a point charge, called the source, over all of space. The inputs for fundamental solutions (Green's functions) are all of space (apart from a singular region), and the output is the "force" exerted from the point throughout space. Green's functions are fundamental to the study of inhomogeneous partial differential equations and are powerful in that they provide a mechanism for obtaining their solutions.

We investigate fundamental solutions of linear partial differential equations on symmetric Riemannian manifolds (harmonic, rank-one symmetric spaces) such as real, complex, quarternionic, and octonionic Euclidean, hyperbolic, and projective spaces. Our recent focus has been on applications of fundamental solutions for linear elliptic partial differential operators on spaces of constant curvature. With G. Olafsson, we are also preparing work on fundamental solutions for the Laplace-Beltrami operator on rank one symmetric spaces of compact and noncompact type [2]. Cohl is preparing a paper with Hirtenstein and Lawrence on deriving Gegenbauer expansions and addition theorems for binomial and logarithmic fundamental solutions of the polyharmonic equation in even-dimensional Euclidean space with powers of the Laplacian greater than or equal to the dimension divided by two [2].

In the following works, we derive and use properties of hypergeometric and basic hypergeometric orthogonal polynomials to derive their properties and properties of the hypergeometric and basic hypergeometric functions themselves. We compute generalizations of generalized and basic hypergeometric orthogonal polynomial generating functions as well as corresponding definite integrals using orthogonality.

We are also interested in the fundamental transformation, representation (symmetry) properties of the special functions and orthogonal polynomials which one often encounters in applied mathematics and mathematical physics. In this regard, in conjunction with J. Park and H. Volkmer, we have submitted a paper which computes all Gauss hypergeometric representations of the Ferrers function of the second kind by starting with the full list of 18 Gauss hypergeometric representations for the associated Legendre function of the second kind [3].

Working with R. Costas-Santos, we have been working on a series of papers which describe the full transformation and representation theory of symmetric basic hypergeometric orthogonal polynomials, namely the Askey-Wilson polynomials (4 symmetric free parameters (sfp)) and its symmetric subfamilies, the continuous dual q-Hahn polynomials (3 sfp), the Al-Salam Chihara polynomials (2 sfp), the continuous big q-Hermite polynomials and the continuous q-Hermite polynomials, and their q-inverse analogues. With Costas-Santos and Ge, we derived all terminating basic hypergeometric representation and transformation properties of the Askey-Wilson polynomials [4]. In a follow-up paper with Costas-Santos, we examine the transformation and representation properties of the Askey-Wilson polynomials with a focus on the terminating very-well poised ${}_8W_7$ representations. We examine the representation relation to the order of the well-known symmetry group of the terminating ${}_4\phi_3$ Askey-Wilson representations given by the symmetric group S_6 in relation to their inversion symmetries and equivalence class collections [5]. With Costas-Santo and Ge we are also finishing up an analogous analysis on the symmetric and q-inverse symmetric sub-families of the Askey-Wilson polynomials [6].

With M. Ismail, we derive 5-term contiguous relations for linearization coefficients of generalized and basic hypergeometric orthogonal polynomials such as Laguerre, Gegenbauer, Hermite, Jacobi, continuous q-ultraspherical/Rogers, and continuous q-Jacobi polynomials [7].

We have also initiated a project which we will continue during the summer with a NIST SURF student to determine all representations of symmetric elliptic integrals and related functions in terms of the respective incomplete Legendre elliptic integrals of the first, second and third kind.

In an ongoing project which focuses on generalized hypergeometric and basic hypergeometric analysis of generating functions and linearization formulae for orthogonal polynomials, we have several ongoing projects. With Costas-Santos, Hwang and Wakhare, our

series-rearrangement technique is extended to generalizations of other generating functions for basic hypergeometric orthogonal polynomials in [8]. Here, we derive generalizations of generating functions for Askey-Wilson, q-ultraspherical/Rogers, q-Laguerre, and little q-Laguerre/Wall polynomials. With Costas-Santos and Zhao, we derive generalized linearization formulas for generalized and basic hypergeometric orthogonal polynomials by applying connection relations to them in [9]. Here, we generalize linearization formulae for continuous q-ultraspherical/Rogers, Jacobi, and continuous q-Jacobi polynomials, Gegenbauer and Laguerre polynomials. With Costas-Santos and Koornwinder, we are investigating dual addition theorems, product formulas and other related formulas for the continuous q-Jacobi polynomials [10]. With Costas-Santos, generalizing some classical formulae for these functions, we derived multi-integral representations for the associated Legendre and Ferrers functions of the first and second kind [11].

Cohl served on the Scientific and Local Organizing committees for the Orthogonal Polynomials and Special Functions Summer School 6 (OPSF-S6) workshop which was held on June 17-23, 2016 at the Norbert Wiener Center for Harmonic Analysis and Applications at the University of Maryland. The OPSF-S6 lecture notes were published by Cambridge University Press, edited by Cohl and Ismail, and included lecture notes by Duran, Ismail, Koelink, Rosengren, and Zeng [12].

Because of the news that Richard Askey entered hospice care in 2019, with 63 colleagues, Cohl and Ismail prepared a *Liber Amicorum for Dick Askey* which was presented to him and his family at an event held on September 4, 2019 at his hometown in Madison, Wisconsin. Unfortunately, Askey passed away on October 9, 2019. Cohl and Ismail are currently preparing an extended version of the Liber Amicorum which will be printed and sent to members of Dick's family and his very close colleagues. It will be composed of the remembrances of 83 of his colleagues. The Liber will be published online with *Celebratio Mathematica* [13]. Cohl, Ismail, and Wu have also submitted a memorial article for Dick Askey to *The Notices of the American Mathematical Society* [14].

Cohl remains editor or co-editor for a special issue on symmetry in special functions and orthogonal polynomials in the journal *Symmetry*; a special volume dedicated to the legacy of Dick Askey for *The Ramanujan Journal*; OP-SF NET of the SIAM Activity Group on Orthogonal Polynomials and Special Functions; and *The Ramanujan Journal*.

[1] G. Olafsson and H. S. Cohl. Fundamental Solutions for the Laplace-Beltrami Operator on the Rank One Symmetric Spaces. In preparation.

[2] H. S. Cohl and J. E. Hirtenstein. Binomial and Logarithmic Gegenbauer Expansions for the Even-dimensional Polyharmonic Equation. In preparation.

[3] H. S. Cohl, J. Park, and H. Volkmer. Gauss Hypergeometric Representations of Ferrers Functions of the Second Kind. In review.

[4] H. S. Cohl, R. S. Costas-Santos and L. Ge. Terminating Basic Hypergeometric Representations and Transformations for the Askey-Wilson Polynomials *Symmetry* **12**:8 (2020), 1290.

[5] H. S. Cohl and R. S. Costas-Santos. Symmetry of Terminating Basic Hypergeometric Representations of the Askey-Wilson Polynomials. In review.

[6] H. S. Cohl, R. S. Costas-Santos and L. Ge. Basic Hypergeometric Transformations from Symmetric and q-inverse Sub-families of the Askey-Wilson polynomials in the q-Askey-scheme. In preparation.

[7] H. S. Cohl and M. E. H. Ismail. Two-dimensional Contiguous Relations for Linearization Formulae. In preparation.

[8] H. S. Cohl, R. S. Costas-Santos, P. R. Hwang, and T. V. Wakhare. Generalizations of Generating Functions for Basic Hypergeometric Orthogonal Polynomials. In review.

[9] H. S. Cohl, R. S. Costas-Santos, and J. Zhao. Generalizations of Linearization Formulae for Continuous Hypergeometric Orthogonal Polynomials. In preparation.

[10] H. S. Cohl, R. S. Costas-Santos, and T. H. Koornwinder. Linearization, Dual Addition Theorem and Product Formulas for the Continuous q-Jacobi Polynomials. In preparation.

[11] H. S. Cohl and R. S. Costas-Santos. Multi-Integral Representations for Associated Legendre and Ferrers Functions. *Symmetry* **12**:10 (2020), 1598, 22 pages.

[12] H. S. Cohl and M. E. H. Ismail, eds. *Lectures on Orthogonal Polynomials and Special Functions, OPSFA Sixth Summer School, Norbert Wiener Center, University of Maryland, College Park, Maryland, July 11-15, 2016.* London Mathematical Society Lecture Note Series **464**, Cambridge University Press, 2020.

[13] H. S. Cohl and M. E. H. Ismail, eds. *Liber Amicorum, a Friendship Book for Dick Askey*. In review.

[14] M. E. H. Ismail, H. S. Cohl, and H.-H. Wu, eds. *The Legacy of Dick Askey (1933-2019)*. In review.

Outreach and Diversity

ACMD staff engage in a variety of efforts that serve to educate the general public about the work of the division and to encourage students to consider careers in science and engineering. We are also involved in internal efforts to improve diversity and inclusivity, which are important for both recrutiment and retention of a high-performing workforce. Some of these efforts are described here.

Gender, Equity and Inclusion Survey Study at the National Institute of Standards and Technology

Justyna P. Zwolak
Mary F. Theofanos (NIST ADLP)
Jasmine Evans (Morgan State University)
Sandra Spickard Prettyman (University of Akron)

In the fall of 2019, NIST funded threes studies to better understand equity and inclusivity at NIST. The purpose of this research effort is to explore the experiences of NIST federal employees to identify the ways in which differential opportunities and outcomes related to gender might exist. The goal is to provide the organization with data about how inequities manifest themselves to inform discussions of policies that foster equitable and inclusive experiences for all employees.

This project focuses specifically on gender and inclusivity and was carried out in three phases. In the first phase, about 10 years of human resource data was analyzed to compare demographics such as positions, promotions, and awards of men and women at NIST. The results of this analysis informed the development of an interview protocol for the second phase of the project. The in-depth interviews with 40 employees from a range of positions and career paths, in turn, informed the development of an organization-wide quantitative online survey of federal employees.

The main research questions addressed in the survey are as follows.

1. What are the differences, if any, in the ways in which men and women experience work at NIST?
2. What, if any, gender inequities exist at NIST? To what extent, if at all, do employees at NIST perceive that gender inequities exist?
3. In what ways, if at all, does NIST culture contribute to a lack of inclusivity?

To provide insight on these questions, the survey included seven sections with questions pertaining to the employees' perception of NIST leadership commitment to diversity and inclusivity (4 questions) and NIST culture (8 questions), their experiences at NIST related to gender (11 questions), their perceptions of their work-related interactions at NIST (16 questions), their beliefs about career advancement, work/life balance, and roles at NIST (13 questions), and about their overall sense of satisfaction from working at NIST and their perceived respect, and recognition at NIST (2 questions). The last section included a series of demographic questions.

The survey was disseminated via email in mid-August 2020 and remained opened for a period of three weeks. More than 1 500 employees responded to the survey for an overall response rate of 33.2 %, with 1 108 completing it. Survey participants' demographics were representative of the overall distribution of staff across directorates, operating units, age, years of service, career paths, pay bands, and educational levels.

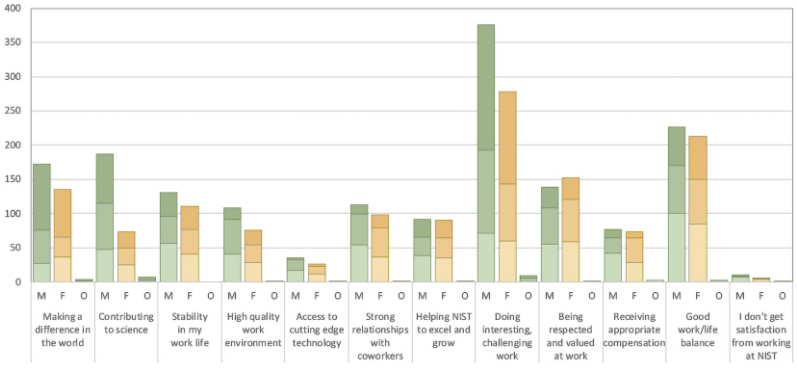


Figure 86. *Results of a sample survey question: Satisfaction from working at NIST. For each question, the responses are divided by gender (M denotes male, F-female, and O-other gender). The respondents were asked to rank their three top choices. The intensity of color represents the rank order, with the darkest color corresponding to rank 1.*

The survey confirms, clarifies and/or expands on the gender related issues identified in Phase 2 of the study, the interviews. The comparison of responses between genders revealed many questions where women and men experience NIST in the same way. For example, women and men agree that teamwork is valued and rewarded at NIST and they believe they "fit in" with colleagues they work with, that opportunities to be hired do not depend on gender and that their need to balance work and other life obligations is supported. Both genders also agree that doing interesting, challenging work gives them a sense of accomplishment that provides the most satisfaction.

However, for a number of questions significant differences were found between genders, providing evidence that men and women have different views and understanding of how gender affects certain aspects of the culture, beliefs, and even interactions at NIST. In particular, statistically significant differences ($p < 0.001$) were found for men and women with respect to the diversity and inclusivity, meritocracy, gendered experiences, promotions, and opportunities.

Overall, the survey results align with the qualitative results and provide quantitative data on the differences in which men and women experience the culture, diversity and inclusivity of NIST. A report summarizing the analysis of the survey data concludes the third and final phase of the project [1].

[1] M. F. Theofanos, J. Evans, J. P. Zwolak, S. Spickard Prettyman. *Survey on Gender, Equity and Inclusion*. NIST Internal Report. In preparation.

Mapping and Analyzing Employee Networks Through the NIST Interactions Survey

Justyna P. Zwolak
Laura Espinal (NIST ADLP)
Camila Young (NIST EL)

In the fall of 2019, the NIST funded three data-driven initiatives to better understand equity and inclusivity. The purpose of the research effort we report on here is to explore the inclusivity of the NIST workforce, including federal employees and associates, through the social network analysis. The goal is to provide the organization with insights about the current climate at NIST as well as suggest strategies that will help promote gender and minority equity and inclusion.

One way to assess inclusivity in a work environment is through the observance of the individuals' direct interactions. Social network analysis (SNA) provides a set of tools suitable to analyze the interaction data which is relational in nature. With the focus of this project being inclusivity at the workplace, we considered two types of relationships: (1) interactions related to achieving work-related goals at NIST, and (2) advice about career-related decisions at NIST. We use the measure of composition of employees' self-reported *ego networks*, that is networks involving only the responding individual and their direct connections, as a proxy for inclusivity. Importantly, while the population of interest is NIST federal employees, to ensure that the reported networks are true representations of the NIST environment, we opted to define the network boundaries by the status of being either a NIST federal employee or an associate. In other words, while the survey was only sent out to NIST federal employees, the respondents were allowed to include NIST associates as their connections.

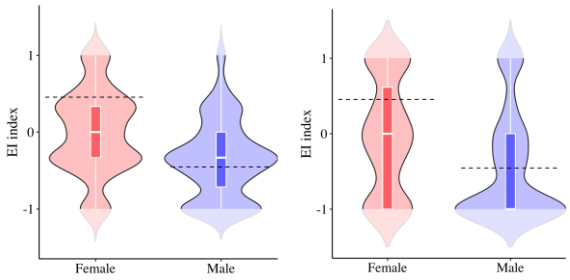


Figure 87. *The distribution of the homophily index between genders. Left: violin plots with overlaid box plots for the homophily index distribution for the advice network. Right: the advice networks. The dashed lines represent theoretical reference point based on the population.*

The NIST Interactions Survey was disseminated via email in early-November 2020 and remained opened for a period of three weeks. In addition to the two network questions, the respondents were asked seven demographic and work history questions about themselves (gender, race, age, supervisory status, tenure, location, and organizational unit (OU)) as well as five questions about each reported connection (gender, relative age (compared to self), supervisory status, location, and OU). More than 800 employees completed the survey for an overall response rate of 24.6 %. Survey participant demographics were fairly representative of the overall distribution of NIST federal employees.

The ego network analysis was carried out in a series of sequential steps, with results from each phase informing the focus of the consecutive phase. In the first stage, we focused on the network's overall structure and size. No statistically significant differences were found when comparing between genders and ethnic groups for either of the network type. Next, we analyzed the ego network compositions. While for the work-related network the majority of differences in composition was correlated with the respondent's location and OU, for the advice network we found additional statistically significant differences in composition based on gender, age group, and tenure. The lack of gender- and ethnicity-based differences suggest that the work-related networks can be thought of as a "network of convenience," capturing the snapshot of current research priorities at NIST. The addition of statistically significant differences in the advice network composition based on gender and age suggest more a preferential nature of these networks, with the connections potentially being actively sought after. This hypothesis was confirmed in the third phase of analysis focusing on the network homophily, that is a measure of similarity between the respondents and their connections. Here, we found that while men's advice networks tens to include mostly men, women's networks were more balanced. However, we also found that a number of women reported networks that include only women,

which given the population of NIST suggest that these women must have made a conscious effort to establish such networks. As such, the advice-related network can be thought of as a "network of choice". A report summarizing the analysis of the NIST Interactions Survey is currently being prepared [1].

[1] L. Espinal, C. Young, and J. P. Zwolak. Mapping and Analyzing Employee Networks through the NIST Interactions Survey. NIST Internal Report. In preparation.

Student Internships in ACMD

Ronald Boisvert

ACMD is committed to helping to prepare the next generation of scientific researchers by providing internships of various types to students at each of the graduate, undergraduate, and high school levels. The NIST programs used to enable such internships include the following:

- *Foreign Guest Researcher Program.* Provides stipends to support visits of guest researchers from foreign institutions for periods of a few weeks to several years.
- *Pathways Program.* Provides temporary Federal appointments to students, typically 1-2 years. Allows easy conversion to full-time permanent status. (Restricted to US Citizens.)
- *Professional Research Experience Program (PREP)*[44]. A cooperative agreement with nine universities[45] that provides a mechanism for NIST to support internships for students from those institutions on the Gaithersburg campus throughout the year. A similar agreement with four universities[46] exists for the NIST Boulder Labs.
- *Student Volunteer Program.* A mechanism that provides unpaid internships for students.
- *Summer High School Internship (SHIP) Program*[47]. SHIP uses the Student Volunteer Program to organize a competitive summer volunteer program for high school students.
- *Summer Undergraduate Research Fellowship (SURF) Program*[48]. A competitive program providing undergraduates a 10-week research experience at NIST.

Funding for all of these programs comes from the Division hosting the student. The Pathways Program, the PREP Program, and the Foreign Guest Researcher Program can also be used to support postdoctoral researchers.

Unfortunately, due to the COVID-19 pandemic, the 2020 SHIP and SURF programs were cancelled. Nevertheless, we still managed to stay engaged with a large number of students. During the last 15 months, ACMD supported the work of 23 student interns, including 19 graduate students and 4 undergraduates. See Table 5 for a complete listing.

ACMD staff members are also active in the education of graduate students, serving both as Ph.D. advisers and as members of thesis committees. See page 149.

[44] https://www.nist.gov/iaao/academic-affairs-office/nist-professional-research-experience-program-prep

[45] Brown University, Georgetown University, Montgomery College, Towson University, the University of the District of Columbia, the University of Maryland College Park, and a consortium of Johns Hopkins University, Morgan State University, and the State University of New York at Binghamton.

[46] Brown University, the Colorado School of Mines, the University of Colorado Boulder, and the University of Colorado Denver

[47] https://www.nist.gov/careers/student-opportunities/summer-high-school-intern-program

[48] https://www.nist.gov/surf

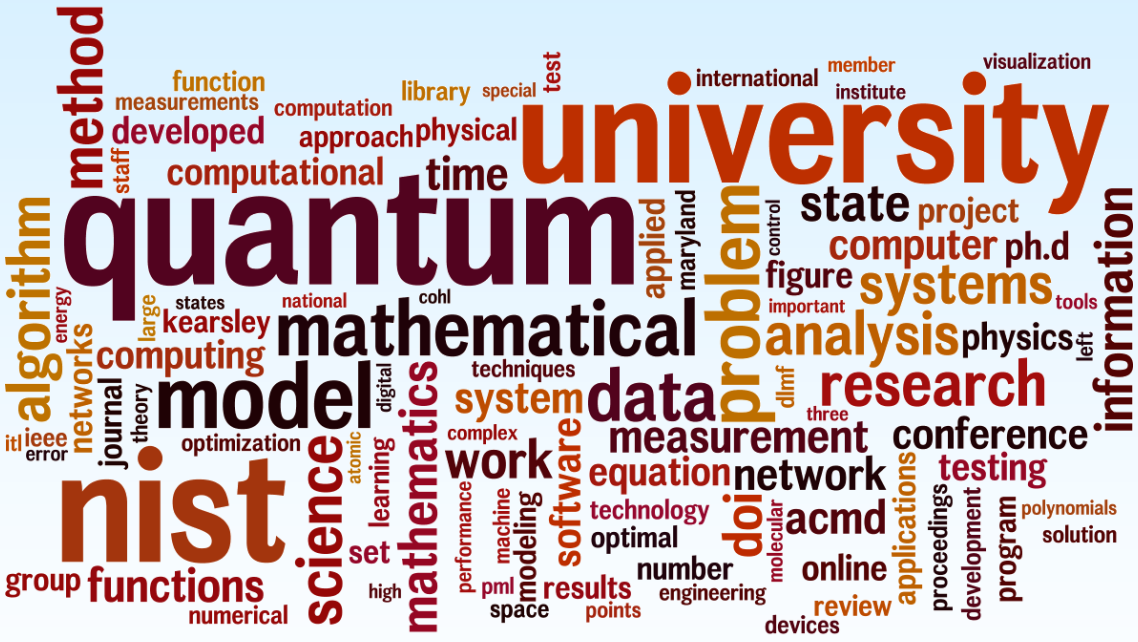
Table 5. *Student interns in ACMD.*

Name	From	Program		Mentor	Topic
Abawonse, Olakunle	Binghamton U.	G	NSF	G. Dogan	Image Analysis in Complex Applications
Agrawal, Sristy	U. Colorado Boulder	G	PREP	S. Glancy	Resource based quantum information theory
Alhejji, Mohammad	U. Colorado Boulder	G	PREP	E. Knill	Quantum randomness protocols
Ariori, Rukayat	Morgan State U.	G	PREP	S. Satterfield	High end visualization
Avagyan, Arik	U. Colorado Boulder	G	PREP	E. Knill	Quantum information processing
Conerly, Azizah	Morgan State U.	U	PREP	J. Kauffman	Numerical turbulence modeling
Cong, Alyssa	Brown University	U	DGR	B. Cloteaux	Rewiring systems under infection avoidance
De Leon, Daniel	U. of Maryland	U	DGR	A. Kearsley	Fractals and Lagrange multiplier estimates
Emiyah, Christian	Morgan State U.	G	PREP	B. Cloteaux	Modeling of contact tracing
Eruslu, Hasan	U. of Delaware	G	FGR	G. Doğan	Algorithms 3D image and shape analysis
Fathima, Hashmath	Morgan State U.	G	PREP	A. Kearsley	Artificial Intelligence for Mass Spectrometry
Geller, Shawn	U. of Colorado	G	PREP	E. Knill	Characterization of quantum state preparation and measurement errors
Gill, Jared	U. of Maryland	U	PREP	J. Terrill	Virtual reality software development
Hu, Mingyu	U. Colorado Boulder	G	PREP	M. Donahue	Modeling of antiferromagnets
Khrac, Katjana	U. Zagreb, Croatia	G	FGR	K. Sayrafian	Positioning metrology in the human body
Kwiatkowski, Alexander	U. of Colorado	G	PREP	E. Knill	Quantum networking
Lauria, Michael	U. of Colorado	G	PREP	E. Knill	Characterization of quantum computers
Mayer, Karl	U. Colorado Boulder	G	PREP	E. Knill	Protocols for quantum optics
Ornstein, Joel	U. of Colorado	G	PREP	S. Glancy	Quantum information journal club
Schneider, Ryan	U. C. San Diego	G	NSF	B. Schneider	Collocation methods to solve the electronic Schrödinger equation
Seilert, Julia	Tech. U. Berlin	G	FGR	A. Kearsley	Parameter identification and optimization in biochemical models
Van Meter, James	U. Colorado Boulder	G	PREP	E. Knill	Quantum measurement of space-time
Wiesing, Tom	Jacobs U. Bremen	G	FGR	B. Miller	Digital Library of Math Functions

Legend
G Graduate Student
U Undergraduate
PREP Professional Research Experience Program
FGR Foreign Guest Researcher
DGR Domestic Guest Researcher
NSF NSF Mathematical Sciences Graduate Internship Program

Part IV

Activity Data



Publications

Note: Names of (co-)authors with a Division affiliation during this reporting period are underlined.

Appeared

Refereed Journals

1. F. Ahmadi, Y. Sozer, M. Donahue, and I. Tsukerman. Low-loss and Lightweight Magnetic Material for Electrical Machinery. *IET Electric Power Applications*, **14** (2020), 282–290. DOI: 10.1049/iet-epa.2019.0430
2. F. Ahmadi, M. Donahue, Y. Sozer, and I. Tsukerman. Micromagnetic Study of Magnetic Nanowires. *AIP Advances* **9** (2019), 125047. DOI: 10.1063/1.5130157
3. D. Anderson, J. Benson, and A. Kearsley. Foundations of Modeling in Cryobiology II: Heat and Mass Transport in Bulk and at Cell Membrane and Ice-Liquid Interfaces. *Cryobiology* **91** (2019), 3-17. DOI: 10.1016/j.cryobiol.2019.09.014
4. D. Anderson, J. Benson, and A. Kearsley. Foundations of Modeling in Cryobiology III: Inward Solidification of a Ternary Solution Towards a Permeable Spherical Cell in the Dilute Limit. *Cryobiology* **92** (2020), 34-46. DOI: 10.1016/j.cryobiol.2019.09.013
5. D. M. Anderson, J. D. Benson, and A. J. Kearsley. Numerical Solution of Inward Solidification of a Dilute Ternary Solution Towards a Semi-permeable Spherical Cell. *Mathematical Biosciences* **316** (2019), 108240. DOI: 10.1016/j.mbs.2019.108240
6. C. Agarwal, J. Klobusicky, and D. Schonfeld. Convergence of Backpropagation with Momentum for Network Architectures with Skip Connections. *Journal of Computational Mathematics* **39** (2021), 147-158. DOI: 10.4208/jcm.1912-m2018-0279
7. J. Arrazola, V. Bergholm, K. Bradler, T. Bromley, M. Collins, I. Dhand, A. Fumagalli, T. Gerrits, A. Goussev, L.Helt, J. Hundal, T. Isacsson, R. Israel, J. Izaac, S. Jahangiri, R. Janik, N. Killoran, S. Kumar, J. Lavoie, A. Lita, D. Mahler, M. Menotti, B. Morrison, S. Nam, L. Neuhaus, H. Y. Qi, N. Quesada, A. Repingon, K. Sabapathy, M. Schuld, D. Su, J. Swinarton, A. Szava, K. Tan, P. Tan, V. Vaidya, Z. Vernon, Z. Zabaneh, and Y. Zhang. Quantum Circuits with Many Photons on a Programmable Nanophotonic Chip. *Nature* **591** (2021), 54. DOI: 10.1038/s41586-021-03202-1
8. S. Asadi, Z. Darvay, G. Lesaja, N. Mahdavi-Amiri, and F. Potra. A Full-Newton Step Interior-Point Method for Monotone Weighted Linear Complementarity Problems. *Journal of Optimization Theory and Applications*, **186** (2020), 864-878. DOI: 10.1007/s10957-020-01728-4
9. D. Becker, B. Alpert, D. Bennett, M. Croce, J. Fowler, J. Gard, A. Hoover, Y. Joe, K. Koehler, J. Mates, G. O'Neil, M. Rabin, C. Reintsema, D. Schmidt, D. Swetz, P. Szypryt, L. Vale, A. Wessels, and J. Ullom. Advances in Analysis of Microcalorimeter Gamma-Ray Spectra. *IEEE Transactions on Nuclear Science* **66**:12 (2019), 2355-2363. DOI: 10.1109/TNS.2019.2953650
10. J. Bernal and J. Lawrence. Characterization and Computation of Matrices of Maximal Trace Over Rotations. *Journal of Geometry and Symmetry in Physics* **53** (2019), 21-53. DOI: 10.7546/jgsp-53-2019-21-53
11. A. Caliciotti, G. Fasano, F. Potra, and M. Roma. Issues on the Use of a Modified Bunch and Kaufman Decomposition for Large Scale Newton's Equation. *Computational Optimization and Applications* **77** (2020), 627-651. DOI: 10.1007/s10589-020-00225-8
12. B. Cloteaux. Forced Edges and Graph Structure. *Journal of Research of NIST* **124** (2019), 22. DOI: 10.6028/jres.124.022
13. H. Cohl, R. Costas-Santos, and L. Ge. Terminating Basic Hypergeometric Representations and Transformations for the Askey-Wilson Polynomials. *Symmetry*. **12**:8 (2020), 1290. DOI: 10.3390/sym12081290
14. H. Cohl and R. Costas-Santos. Multi-Integral Representations for Associated Legendre and Ferrers Functions. *Symmetry*. **12**:10 (2020), 1598. DOI: 10.3390/sym12101598
15. R. Evans, A. Balijepalli, and A. Kearsley. Transport Phenomena in Biological Field Effect Transistors. *SIAM Journal on Applied Mathematics* **80**:6 (2020), 2586-2607. DOI: 10.1137/19M1255495
16. P. Faist, S. Nezami, V. V. Albert, G. Salton, F. Pastawski, P. Hayden, and J. Preskill. Continuous Symmetries and Approximate Quantum Error Correction. *Physical Review X* **10** (2020), 041018. DOI: 10.1103/PhysRevX.10.041018
17. J. Fowler, B. Alpert, Y.-I. Joe, G. O'Neil, D. Swetz, and J. N. Ullom. A Robust Principal Component Analysis for Messy Microcalorimeter Data. *Journal of Low Temperature Physics* **199** (2020), 745-753. DOI: 10.1007/s10909-019-02248-w
18. T. Gerrits, A. Migdall, J. C. Bienfang, J. Lehman, S. W. Nam, J. Splett, I. Vayshenker, and J. Wang. Calibration of Free-space and Fiber-coupled Single-photon Detectors. *Metrologia* **57** (2020). DOI: 10.1088/1681-7575/ab4533

19. L. Ghandehari, Y. Lei, R. Kacker, D. Kuhn, T. Xie, and D. Kung. A Combinatorial Testing-Based Approach to Fault Localization. *IEEE Transactions on Software Engineering*, **46** (2020), 616-645. DOI: 10.1109/TSE.2018.2865935

20. H. Gharibnejad and B. Schneider. Numerical Methods Every Atomic and Molecular Physicist Should Know, *Nature Reviews Physics* **2,** 89–102 (2020). DOI: 10.1038/s42254-019-0126-3

21. Z. Gimbutas, S. Jiang, and L.-S. Luo. Evaluation of Abramowitz Functions in the Right Half of the Complex Plane. *Journal of Computational Physics* **405** (2020), 109169. DOI: 10.1016/j.jcp.2019.109169

22. Z. Gimbutas, N. Marshall, and V. Rokhlin. A Fast Simple Algorithm for Computing the Potential of Charges on a Line. *Applied and Computational Harmonic Analysis* **49**:3 (2020), 815-830. DOI: 10.1016/j.acha.2020.06.002

23. A. Greiner-Petter, A. Youssef, T. Ruas, B. R. Miller, M. Schubotz, A. Aizawa, and B. Gipp. Math-word Embedding in Math Search and Semantic Extraction. *Scientometrics* **125** (2020), 3017-3046. DOI: 10.1007/s11192-020-03502-9

24. L. Hu, W. Wong, D. Kuhn, and R. Kacker. How Does Combinatorial Interaction Testing Perform in the Real World: An Empirical Study on Five Industrial Systems. *Journal of Empirical Software Engineering*, **25** (2020), 2661-2694. DOI: 10.1007/s10664-019-09799-2

25. F. Hunt. A First Hitting Time Approach to Finding Effective Spreaders in a Network. *AMS Contemporary Mathematics Series* **759** (2020), 71-91. DOI: 10.1090/conm/759

26. F. Hunt and R. Pozo. Fast Methods for Finding Multiple Effective Influencers in Real Networks. *Journal of Research of NIST* **125** (2020), 125036. DOI: 10.6028/jres.125.036

27. A. Irfan, K. Mayer, G. Ortiz, and E. Knill. Certified Quantum Measurement of Majorana Fermions. *Physical Review A* **101** (2020), 032106. DOI: 10.1103/PhysRevA.101.032106

28. E. Kim, G. McFadden, and A. Cerfon. Elimination of MHD Current Sheets by Modifications to the Plasma Wall in a Fixed Boundary Model. *Plasma Physics and Controlled Fusion*, **62**:4 (2020) 044002. DOI: 10.1088/1361-6587/ab6d48

29. J. Klobusicky, J. Fricks, and P. Kramer. Effective Behavior of Cooperative and Nonidentical Molecular Motors. *Research in the Mathematical Sciences* **7** (2020), 29. DOI: 10.1007/s40687-020-00230-7

30. J. Klobusicky, G. Menon, and R. Pego. Two-Dimensional Grain Boundary Networks: Stochastic Particle Models and Kinetic Limits. *Archive for Rational Mechanics and Analysis* **239** (2020), 301-355. DOI: 10.1007/s00205-020-01577-5

31. B. Korzh, Q. Zhao, J. Allmaras, S. Frasca, T. Autry, E. Bersin, A. Beyer, R. Briggs, B. Bumble, M. Colangelo, G. Crouch, A. Dane, T. Gerrits, A. Lita, F. Marsili, G. Moody, C. Peña, E. Ramirez, J. Rezac, N. Sinclair, M. Stevens, A. Velasco, V. Verma, E. Wollman, S. Xie, D. Zhu, P. Hale, M. Spiropulu, K. Silverman, R. Mirin, S. Nam, A. Kozorezov, M. Shaw, and K. Berggren. Demonstration of sub-3 ps Temporal Resolution with a Superconducting Nanowire Single-Photon Detector. *Nature Photonics* **14** (2020), 250-255. DOI: 10.1038/s41566-020-0589-x

32. D. Kuhn, R. Kacker, Y. Lei, and D. Simos. Input Space Coverage Matters. *IEEE Computer* **55** (2020), 37-44. DOI: 10.1109/MC.2019.2951980

33. P. S. Kuo, V. B. Verma and S. W. Nam. Demonstration of a polarization-entangled photon-pair source based on phase-modulated PPLN. *OSA Continuum* **3**:2 (2020) 2951. DOI: 10.1364/OSAC.387449

34. J. Lawrence, J. Bernal, and C. Witzgall. A Purely Algebraic Justification of the Kabsch-Umeyama Algorithm. *Journal of Research of NIST* **124** (2019), 124028. DOI: 10.6028/jres.124.028

35. S. Lieu, R. Belyansky, J. T. Young, R. Lundgren, V. V. Albert, and A. V. Gorshkov. Symmetry Breaking and Error Correction in Open Quantum Systems. *Physical Review Letters* **125** (2020), 240405. DOI: 10.1103/PhysRevLett.125.240405

36. L. Ma, X. Tang and O. Slattery. Optical Quantum Memory and its Applications in Quantum Communication Systems. *Journal of Research of NIST* **125** (2020), 125002. DOI: 10.6028/jres.125.002

37. J. Majikes, P. Patrone, D. Schiffels, M. Zwolak, A. Kearsley, S. Forry, and J. Liddle. Revealing Thermodynamics of DNA Origami Folding via Affine Transformations. *Nucleic Acids Research* **48** (2020), 5268-5280. DOI: 10.1093/nar/gkaa283

38. J. Majikes, P. N. Patrone, A. J. Kearsley, M. Zwolak, and J. A. Liddle. Failure Mechanisms in DNA Self-assembly: Barriers to Single Fold Yield. *ACS Nano* **15** (2021) 3284–3294.

39. N. Martys, W. George, R. Murphy, and K. Weigandt. Pipe Flow of Sphere Suspensions Having a Power-Law-Dependent Fluid Matrix. *Journal of Rheology* **64** (2020), 445. DOI: 10.1122/1.5131021

40. G. McFadden, W. Boettinger, and Y. Mishin. Effect of Vacancy Creation and Annihilation on Grain Boundary Motion. *Acta Materialia* **185** (2020), 66-79. DOI: 10.1016/j.actamat.2019.11.044

41. A. Moorthy, A. Kearsley, W. Mallard, and W. Wallace. Mass Spectral Similarity Mapping Applied to Fentanyl Analogs. *Forensic Chemistry* **19** (2020), 100237. DOI: 10.1016/j.forc.2020.100237

42. R. Nehra, M. Eaton, C. González-Arciniegas, M. Kim, T. Gerrits, A. Lita, S. Nam and O. Pfister. Generalized Overlap Quantum State Tomography. *Physical Review Research* **2** (2020) 042002. DOI: 10.1103/PhysRevResearch.2.042002

43. P. Patrone, A. Dienstfrey, and G. McFadden. Model Reduction of Rigid-Body Molecular Dynamics via Generalized Multipole Potentials. *Physical Review E* **100** (2019), 063302. DOI: 10.1103/PhysRevE.100.063302

44. P. Patrone, A. Kearsley, J. Majikes, and J. Liddle. Analysis and Uncertainty Quantification of DNA Fluorescence Melt Data: Applications of Affine Transformations. *Analytical Biochemistry* **607** (2020), 113773. DOI: 10.1016/j.ab.2020.113773

45. P. Patrone, E. Romsos, M. Cleveland, P. Vallone, and A. Kearsley. Affine Analysis for Quantitative PCR Measurements. *Analytical and Bioanalytical Chemistry* **412** (2020), 7977-7988. DOI: 10.1007/s00216-020-02930-z

46. P. Patrone, A. Li, G. Cooksey, and A. Kearsley. Measuring Microfluidic Flow Rates: Monotonicity, Convexity, and Uncertainty. *Applied Mathematics Letters* **112** (2021), 106694. DOI: 10.1016/j.aml.2020.106694

47. B. V. Saunders. Complex Variables, Mesh Generation, and 3D Web Graphics: Research and Technology Behind the Visualizations in the NIST Digital Library of Mathematical Functions. *AMS Contemporary Mathematics Series* **759** (2020), 145-156. DOI: 10.1090/conm/759/15272

48. J. Sims, B. Padhy, and M. Ruiz. Exponentially Correlated Hylleraas-Configuration-Interaction Nonrelativistic Energy of the singlet S Ground State of the Helium Atom. *International Journal of Quantum Chemistry*, (2020), e26470. DOI: 10.1002/qua.26470

49. J. Sims. Hylleraas-Configuration-Interaction Nonrelativistic Energies for the 3 singlet S, 4 singlet S, 5 singlet S, 6 singlet S, and 7 singlet S Excited States of the Beryllium Atom. *Journal of Research of NIST* **125** (2020), 125006. DOI: 10.6028/jres.125.006

50. J. Sims and M. Ruiz. Parallel Generalized Real Symmetric-Definite Eigenvalue Problem. *Journal of Research of NIST* **125** (2020), 125032. DOI: 10.6028/jres.125.032

51. G. Thekkadath, M. Mycroft, B. Bell, C. Wade, A. Eckstein, D. Phillips, R. Patel, A. Buraczewski, A. Lita, T. Gerrits, S. Nam, M. Stobińska, A. Lvovsky and I. Walmsley. Quantum-Enhanced Interferometry with Large Heralded Photon-Number States. *npj Quantum Information* **6** (2020), 89. DOI: 10.1038/s41534-020-00320-y

52. L. Tian, E. Elsheikh, P. N. Patrone, A. J. Kearsley, A. Gaigalas, S. Inwood, S. Lin-Gibson, D. Esposito, and L. Wang. Towards Quantitative and Standardized Serological and Neutralization Assays for COVID-19. *International Journal of Molecular Sciences* **22**:5 (2021), 2723. DOI: 10.3390/ijms22052723

53. C. Titus, D. Li, B. Alpert, H.-M. Cho, J. Fowler, S.-J. Lee, K. Morgan, D. Swetz, J. Ullom, A. Wessels, and K. Irwin. Count Rate Optimizations for TES Detectors at a Femtosecond X-ray Laser. *Journal of Low Temperature Physics* **199** (2020), 1038-1045. DOI: 10.1007/s10909-020-02379-5

54. L. Wang, R. Bhardwaj, H. Mostowski, P. Patrone, A. J. Kearsley, J. Watson, L. Lim, J. Pichaandi, O. Ornatsky, D. Majonis, S. Bauer, H. Degheidy. Establishing B-cell Reference Control Materials for Comparable and Quantitative Cytometric Expression Analysis. *PLOS One* **16**:3 (2021), 1-18. DOI: 10.1371/journal.pone.0248118

55. E. Williams, J. Zwolak, R. Dou, and E. Brewe. Linking Engagement and Performance: The Social Network Analysis Perspective. *Physical Review Physics Education Research* **15** (2019), 020150. DOI: 10.1103/PhysRevPhysEducRes.15.020150

56. Y. Zhang, H. Fu, and E. Knill. Efficient Randomness Certification by Quantum Probability Estimation. *Physical Review Research* **2** (2020), 013016. DOI: 10.1103/PhysRevResearch.2.013016

57. Y. Zhang, L. Shalm, J. Bienfang, M. Stevens, M. Mazurek, S. Nam, C. Abellán, W. Amaya, M. Mitchell, H. Fu, C. Miller, A. Mink, and E. Knill. Experimental Low-Latency Device-Independent Quantum Randomness. *Physical Review Letters* **124** (2020), 010505. DOI: 10.1103/PhysRevLett.124.010505

58. J. Zwolak, T. McJunkin, S. Kalantre, J. Dodson, E. MacQuarrie, D. Savage, M. Lagally, S. Coppersmith, M. Eriksson, and J. Taylor. Auto-Tuning of Double Dot Devices *in situ* with Machine Learning. *Physical Review Applied* **13** (2020), 034075. DOI: 10.1103/PhysRevApplied.13.034075

Books

1. C. Benzmüller and B. Miller (eds.). *Intelligent Computer Mathematics,* Proceedings of the 13th International Conference (CICM 2020), *Lecture Notes in Artificial Intelligence* **12236**, Springer, 2020. DOI: 10.1007/978-3-030-53518-6

2. M. Donahue (ed.). *Electrostatic and Magnetic Phenomena*. World Scientific, 2020. DOI: 10.1142/10987-vol1

3. H. Cohl and M. Ismail (eds.). *Lectures on Orthogonal Polynomials and Special Functions*. London Mathematical Society Lecture Note Series **464**, Cambridge University Press, 2020. DOI: 10.1017/9781108908993

4. J. Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Verlag, 2020. DOI: 10.1007/978-3-030-52915-4

Book Chapters

1. D. Porter and M. Donahue. Standard Problems in Micromagnetics. *Electrostatic and Magnetic Phenomena* (M. Donahue, ed.), World Scientific, 2020, 285–324. DOI: 10.1142/9789813270268_0009

In Conference Proceedings

1. S. Bhushan, O. Slattery, X. Tang and L. Ma. Terahertz Electromagnetically Induced Transparency in Cesium Atoms. In *Frontiers in Optics / Laser Science*, (B. Lee, C. Mazzali, K. Corwin, and R. Jason Jones, eds.), OSA Technical Digest, Optical Society of America, September 14-17, 2020, JTu1B.40. DOI: 10.1364/FIO.2020.JTu1B.40

2. J. Chandrasekaran, H. Feng, Y. Lei, R. Kacker, and D. Kuhn. Effectiveness of Dataset Reduction in Testing Machine Learning Algorithms. In *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence Testing* (AITest), Online, August 3-6, 2020, 133-140. DOI: 10.1109/AITEST49225.2020.00027

3. M. Coudron and S. Menda. Computations with Greater Quantum Depth Are Strictly More Powerful (Relative to an Oracle). In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2020),* Online, June 2020, 889–901. DOI: 10.1145/3357713.3384269

4. T. Gerrits, A. Migdall, J. Bienfang, J. Lehman, S. Nam, O. Slattery, J. Spleet, I. Vayshenker and C. Wang. Calibration of Free-Space and Fiber-Coupled Single Photon Detectors. In *Proceedings of the OSA Conference on Quantum 2.0*, Online, September 14-17, 2020, QW6B.7. DOI: 10.1364/QUANTUM.2020.QW6B.7

5. D. Ginev and B. Miller. Scientific Statement Classification over arXiv.org. In *Proceedings of the 12th Language Resources and Evaluation Conference*, Online, May 2020, 1219-1226. URL: https://www.aclweb.org/anthology/2020.lrec-1.153/

6. A. Greiner-Petter, M. Schubotz, F. Müller, C. Breitinger, H. Cohl, A. Aizawa, and B. Gipp. Discovering Mathematical Objects of Interest-A Study of Mathematical Notations. In *Proceedings of the Web Conference (WWW'20)*, Online, April 2020, 1445-1456. DOI: 10.1145/3366423.3380218

7. R. Kacker, R. Kessel, and K. Sommer. Operational Measurement Uncertainty and Bayesian Probability Distribution. In, *Proceedings of the Conference: Sensor and Measurement Science International* (SMSI 2020) *Measurement Science D3.1*, Online, June 22-25, 2020, 275-276. DOI: 10.5162/SMSI2020/D3.1

8. D. Kuhn, R. Kacker, Y. Lei, and D. Simos. Combinatorial Methods for Explainable AI. In *Proceedings of the 2020 IEEE International Conference on Software Testing, Verification and Validation Workshops* (ICSTW), Online, October 24-28, 2020, 167-170. DOI: 10.1109/ICSTW50294.2020.00037

9. P. S. Kuo, Effect of Amplitude Mismatch on Entanglement Visibility in Photon-Pair Sources. In *OSA Advanced Photonics Congress 2020 OSA Technical Digest*, Online, July 13-16, 2020. DOI: 10.1364/IPRSN.2020.ITu4A.15

10. R. La. Identifying Vulnerable Nodes to Cascading Failures: Optimization-Based Approach. In *Proceedings of the 9th International Conference on Complex Networks and Their Applications* (Complex Networks '19), Lisbon, Portugal, December 1-3, 2019. DOI: 10.1007/978-3-030-36687-2_64

11. V. Mai, R. La, and A. Battou. Optimal Cybersecurity Investments for SIS Model. In *Proceedings of the 2020 IEEE Globecom Communications Conference (GLOBECOM'2020)*, Online, December 7-11, 2020. DOI: 10.1109/GLOBECOM42002.2020.9348109

12. V. Marbukh. Towards Scalable Optimization of Large-Scale IoT Access for Mobile Users with Diverse Requirements. In *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (PIMRC'2020), Online, August 31 - September 3, 2020. DOI: 10.1109/PIMRC48278.2020.9217226

13. A. Mikhaylov, K. Parzuchowski, M. Mazurek, R. Wilson, T. Gerrits, D. Lum, C. Camp, M. Stevens, and R. Jimenez. Setting Limits on Two-Photon Absorption Cross Sections in Common Fluorescent Molecules with Entangled Photon Pairs Excitation. In *Proceedings of the 2020 Conference on Lasers and Electro-Optics*, Online, May 10-15, 2020. DOI: 10.1364/CLEO_AT.2020.JTh3N.4

14. A. S. Moorthy and A. J. Kearsley. Pattern Similarity Measures Applied to Mass Spectra. In *Progress in*

Figure 88. *(left to right) Greg Cooksey (NIST PML), Paul Patrone (ACMD), and Lili Wang (MML) describe NIST research in cytometry to visitors Heba Degheidy and Steven Bauer of the FDA. For details on this project, see page 52.*

Industrial Mathematics: Success Stories: The Industry and the Academia Points of View (M. Cruz, C. Parés, and P. Quintela, eds.), ICIAM 2019, Valencia Spain. 43-54.

15. O. Narayan, I. Saniee, and V. Marbukh. Congestion Due to Random Walk Routing. In *Proceedings of the 9th International Conference on Complex Networks and Their Applications* (ComplexNetworks 2020), Online, December 1-3, 2020. DOI: 10.1007/978-3-030-65347-7_46

16. M. Raunak, D. Kuhn, R. Kogut, and R. Kacker. Vulnerability Trends in Web Servers and Browsers, In *Proceedings of the Symposium on the Science of Security* (HotSoS), Online, September 22-23, 2020. DOI: 10.1145/3384217.3384227

17. M. Roudneshin, K. Sayrafian, and A. G. Aghdam. A Machine Learning Approach to the Estimation of Near-Optimal Electrostatic Force in Micro Energy-Harvesters. In *Proceedings of the IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE'19)*, Ottawa, Canada, October 16-18, 2019. DOI: 10.1109/WiSEE.2019.8920332

18. B. Rouzbehani, V. Marbukh, and K. Sayrafian. A Joint Admission Control & Resource Management Scheme for Virtualized Radio Access Networks. In *Proceedings of the 5th Annual IEEE Conference on Standards for Communications and Networking (CSCN'2019)*, Granada, Spain, October 28-30, 2019. DOI: 10.1109/CSCN.2019.8931373

19. B. Rouzbehani, V. Marbukh, and K. Sayrafian. Utility Proportional Resource Allocation for Users with Diverse SLAs in Virtualized Radio Access Networks. In *Proceedings of the IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom'2020)*, Online, May 26-29, 2020. DOI: 10.1109/BlackSeaCom48709.2020.9235011

20. M. Roudneshin, K. Sayrafian, and A. G. Aghdam. Adaptive Estimation of Near-Optimal Electrostatic Force in Micro Energy-Harvesters. In *Proceedings of the IEEE Conference on Control Technology and Applications (CCTA)*, Online, August 24-26, 2020. DOI: 10.1109/CCTA41146.2020.9206354

21. P. Scharpf, M. Schubotz, A. Youssef, F. Hamborg, N. Meuschke, and B. Gipp. Classification and Clustering of arXiv Documents, Sections, and Abstracts, Comparing Encodings of Natural and Mathematical Language. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, JCDL 2020, Online, August 2020, 137-146. DOI: 10.1145/3383583.3398529

22. B. Schneider, K. Bartschat, O. Zatsarinny, K. Hamilton, I. Bray, A. Scrinzi, F. Martin, J. Vasquez, J. Tennyson, J. Gorfinkiel, R. Lucchese and S. Pamidighantam. Atomic and Molecular Scattering Applications in an Apache Airavata Science Gateway. In *Proceedings of PEARC '20: Practice and Experience in Advanced Research Computing*, Online, July 2020, 270-277. DOI: 10.1145/3311790.3397342

23. M. Wagner, K. Kleine, D. Simos, D. Kuhn, and R. Kacker. CAGEN: A Fast Combinatorial Test Generation Tool with Support for Constraints and Higher-Index Arrays. In *Proceedings of the 2020 IEEE International Conference on Software Testing, Verification and Validation Workshops* (ICSTW-2020), Online, October 24-28, 2020, 191-200. DOI: 10.1109/ICSTW50294.2020.00041

24. A. Youssef and B. Miller. A Contextual and Labeled Math-Dataset Derived from NIST's DLMF. In *Intelligent Computer Mathematics*, Proceedings of the 13th International Conference, CICM 2020, Online, July 26–31, 2020 (C. Benzmüller and B. Miller, eds., *Lecture Notes in Artificial Intelligence* **12236**, Springer, 1219-1226, 2020. DOI: 10.1007/978-3-030-53518-6_25

25. J. Zwolak, S. Kalantre, T. McJunkin, B. Weber, and J. Taylor. A Ray-Based Classification Framework for High-Dimensional Data. In *Proceedings of the Machine Learning and the Physical Sciences Workshop,* 34th Conference on Neural Information Processing Systems (NeurIPS), Online, December 11, 2020. URL: https://ml4physicalsciences.github.io/2020/files/NeurIPS_ML4PS_2020_154.pdf

Technical Reports

1. G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, J. Kelsey, Y.-K. Liu, C. Miller, D. Moody, R. Peralta, R. Perlner, A. Robinson, and D. Smith-Tone. Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process. NISTIR 8309, July 2020. DOI: 10.6028/NIST.IR.8309
2. R. F. Boisvert (ed.). Applied and Computational Mathematics Division: A Summary of Activities in Fiscal Year 2019. NISTIR 8306, April 30, 2020. DOI: 10.6028/NIST.IR.8306
3. B. Cloteaux. How Much Regularity Forces a Sequence to be Graphic?" arXiv preprint 2009.07135, (2020). URL: https://arxiv.org/abs/2009.07135
4. N. J. Coble and M. Coudron. Quasi-polynomial Time Approximation of Output Probabilities of Constant-depth, Geometrically-local Quantum Circuits. arXiv Preprint 2012.05460 (2020). URL: https://arxiv.org/abs/2012.05460
5. S. Kotler, G. Peterson, E. Shojaee, F. Lecocq, K. Cicak, A. Kwiatkowski, S. Geller, S. Glancy, E. Knill, R. Simmonds, J. Aumentado, and J. Teufel. Tomography of Entangled Macroscopic Mechanical Objects, arXiv preprint 2004.05515 (2020). URL: https://arxiv.org/abs/2004.05515
6. A. Olivas, C. Ferraris, N. Martys, W. George, E. Garboczi, and B. Toman. Certification of SRM 2493: Standard Reference Mortar for Rheological Measurements. NISTSP 260-186, October 2017. DOI: 10.6028/NIST.SP.260-187
7. L. Shalm, Y. Zhang, J. Bienfang, C. Schlager, M. Stevens, M. Mazurek, C. Abellán, W. Amaya, M. Mitchell, M. Alhejji, H. Fu, J. Ornstein, R. Mirin, S. Nam, and E. Knill. Device-Independent Randomness Expansion with Entangled Photons, arXiv preprint 1912.11158 (2019). URL: https://arxiv.org/abs/1912.11158
8. B. Schneider, K. Bartschat, O. Zatsarinny, I. Bray, A. Scrinzi, F. Martin, M. Klinker, J. Tennyson, J. Gorfinkiel, and S. Pamidighantam. A Science Gateway for Atomic and Molecular Physics. arXiv preprint 2001.02286. URL: https://arxiv.org/abs/2001.02286

Blog Posts

1. R. F. Boisvert. A Tribute to William F. Mitchell. SIAM News Blog, November 7, 2019. URL: https://sinews.siam.org/Details-Page/a-tribute-to-william-f-mitchell
2. J. P. Zwolak, Ebb and Flow: Creating Quantum Dots Automatically With AI. Taking Measure Blog, NIST, September 4, 2018. URL: https://www.nist.gov/blogs/taking-measure/ebb-and-flow-creating-quantum-dots-automatically-ai

Accepted

1. S. Ambroziak and K. Sayrafian. IoT for Healthcare Applications. In *COST CA15104 Book: Inclusive Radio Communication Networks for 5G and Beyond*, Elsevier.
2. M. Coudron, J. Stark, and T. Vidick. Trading Locality for Time: Certifiable Randomness from Low-Depth Circuits. *Communications in Mathematical Physics*.
3. J. Fowler, G. O'Neil, B. Alpert, D. Bennett, E. Denison, W. Doriese, G. Hilton, L. Hudson, Y.-I. Joe, K. Morgan, D. Schmidt, D. Swetz, C. I. Szabo, and J. Ullom. Absolute Energies and Emission Line Shapes of the L X-ray Transitions of Lanthanide Metals. *Metrologia*.
4. C. Qu, B. I. Schneider, T. Allison, A. J. Kearsley, and W. Keyrouz. Predicting Kováts Retention Indices Using Graph Neural Networks. *Journal of Chromatography A*.
5. Y. Zhai, N. S. Martys, W. L. George, J. Nayem, and Y. Liu. Intermediate Scattering Functions of a Rigid Body Monoclonal Antibody Protein in Solution Studied by Dissipative Particle Dynamic Simulation. *Structural Dynamics*.

In Review

1. J. Bernal, J. Lawrence, G. Doğan, and C. Hagwood. Computing Elastic Shape Distances Between Curves in D-Dimensional Space.
2. A. Carasso. Stabilized Leapfrog Scheme Run Backward in Time, and the Explicit Stepwise Computation of Ill-Posed Time-Reversed 2D Navier-Stokes Equations.
3. B. Cloteaux. How Much Regularity Forces a Sequence to Be Graphic?
4. H. Cohl, J. Park, and H. Volkmer. Gauss Hypergeometric Representations of the Ferrers Function of the Second Kind.
5. H. Cohl and R. Costas-Santos. Symmetry of Terminating Series Representations of the Askey-Wilson Polynomials.
6. H. Cohl, R. Costas-Santos, P. Hwang and T. Wakhare. Generalizations of Generating Functions for Basic Hypergeometric Orthogonal Polynomials.
7. H. S. Cohl, M. Ismail and H.-H. Wu (eds.). The Legacy of Dick Askey (1933-2019).

8. E. Fleisig and G. Doğan. VEMOS: A GUI for Evaluation of Similarity Metrics on Complex Data Sets.
9. A. Greiner-Petter, H. Cohl, A. Youssef, M. Schubotz, A. Trost, R. Dey, A. Aizawa, and B. Gipp. Comparative Verification of the Digital Library of Mathematical Functions and Computer Algebra Systems.
10. A. Kearsley and A. Moorthy. Mathematics and Mass Spectra: Model Problems to Study the Fentanyl Epidemic.
11. V. Mai, R. La, and A. Battou. Optimal Cybersecurity Investments in Large Networks Using SIS Model: Algorithm Design.
12. K. Parzuchowski, A. Mikhaylov, M. Mazurek, R. Wilson, D. Lum, T. Gerrits, C. Camp Jr., M. Stevens, and R. Jimenez. Setting Bounds on Two-Photon Absorption Cross-Sections in Common Fluorophores with Entangled Photon Pair Excitation.
13. P. Patrone, A. Kearsley, E. Romsos, and P. Vallone. Improving Baseline Subtraction for Increased Sensitivity of Quantitative PCR Measurements.
14. C. You, P. Bierhorst, A. Lita, S. Glancy, N. Bhusal, S. Kolthammer, J. Dowling, E. Knill, S. Nam, R. Mirin, O. Magana-Loaiza, and T. Gerrits. Multiphoton Quantum Metrology Without Pre- and Post-Selected Measurements.

Inventions

Patents Awarded

1. R. Kuhn and R. Kacker. Oracle-Free Match Testing of a Program Using Covering Arrays and Equivalence Classes. Patent US 10,552,300B2, February 4, 2020.
2. L. Ma, X. Tang, and O. Slattery. Direct Absolute Spectrometer for Direct Absolute Spectrometry. Patent US 10,641,655, May 5, 2020.

Patents in Review

1. G. Cooksey, A. Kearsley, and P. Patrone. Repeated Single Cell Cytometry in an Optofluidic Chip. Provisional Patent Application, NIST Docket 20-043P1, June 22, 2020.
2. G. Cooksey, A. Kearsley, and P. Patrone. Multiplexed Amplitude Modulation Photometer and Performing Multiplexed Amplitude Modulation Photometry. United States Letters Patent Application, Serial Number 17/084,683, October 30, 2020.
3. A. Kearsley, P. Patrone, E. Romsos, and P. Vallone. Process for Affine qPCR Background Subtraction and Data Analysis. Provisional Patent Application, Docket 20-052p1, June 17, 2020.
4. J. Ullom, G. O'Neil, L. Avila, K. Silverman, D. Swetz, R. Jimenez, W. Doriese, G. Hilton, C. Reintsema, D. Schmidt, B. Alpert, J. Uhlig, Y. Joe, W. Fullagar, V. Sundstrom, I. Maasilta, and J. Fowler. X-RAY SPECTROMETER. US Patent Application, 20190064084.
5. J. Zwolak, J. Taylor, S. Kalantre, and T. McJunkin. Ray-Based Classification Framework for Machine Learning-Based Tuning Techniques for Experiments, NIST Docket #20-059P1, US Provisional Patent Application 63/083,368, September 25, 2020.

Presentations

Note: When multiple presenters are listed, names of co-presenters with a Division affiliation during this reporting period are underlined.

Invited Talks

1. D. Anderson, S. Coriell, P. Guba, G. McFadden, and B. Murray. "Convective Instabilities in Ternary Alloy Solidification." Oxford Center for Industrial and Applied Mathematics Seminar, Oxford University, Online, November 25, 2020.
2. R. F. Boisvert, "Computational Science in a Federal Laboratory." Mathematics Department Colloquium, Morgan State University, October 3, 2019.
3. R. F. Boisvert (Panelist). QuICS 5-Year Anniversary Research Symposium. University of Maryland, College Park, MD, January 22, 2020.
4. R. F. Boisvert, "Reproducibility Badging." International Federation for Information Processing (IFIP) Working Group 2.5 (Numerical Software) Annual Meeting, Online, August 19, 2020.
5. R. F. Boisvert, "Special Functions at NIST: From Mathematical Tables to Digital Libraries." Fisk Colloquium, University of Wyoming, Online, November 13, 2020.
6. A. Cerfon, E. Kim, and G. McFadden. "Progress and Challenges for the Robust Calculation of Equilibrium Configurations in BIEST and NSTAB." Hidden Symmetries and Fusion Energy (Virtual) Annual Team Meeting, Simons Collaboration on Hidden Symmetries and Fusion Energy, Princeton Center for Theoretical Science, Online, August 3-7, 2020.
7. M. Coudron. "Computations with Greater Quantum Depth Are Strictly More Powerful (Relative to an

Oracle).” Mathematics Department, University of Maryland, Online, November 2020.

8. G. Doğan. J. Bernal, and C. Hagwood. “Fast Optimization Algorithms for Shape-based Data Analysis.” SIAM Conference on Mathematical Data Science, Online, May 4 - June 30, 2020.
9. G. Doğan and H. Eruslu. “Algorithms for Variational Segmentation of Regions and Boundaries.” SIAM Conference on Imaging Science, Online, July 6-17, 2020.
10. R. Evans, A. Balijepalli, and A. Kearsley. “Transport Phenomena in Field Effect Transistors.” Applied and Computational Math Seminar, George Mason University, Fairfax, VA, November 1, 2019.
11. T. Gerrits. “Single-Photon Source, Detector and Component Characterization for Future Quantum Networks.” Photonics North Conference, Online, May 26-28, 2020.
12. T. Gerrits. “Single-Photon Source, Detector and Component Characterization for future Quantum Networks”. Quantum Network Grand Challenge Seminar Series, NIST, Online, March 27, 2020.
13. A. Kearsley. “Control of Inward Solidification in Cryobiology.” American University Mathematics and Statistics Colloquium, American University, Online, November 10, 2020.
14. A. Kearsley. “A Survey of Applied Mathematics Applications in Metrology at NIST.” Mathematical Association of America, MD-DC-VA Section, Online, November 7, 2020.
15. E. Kim, G. McFadden, and A. Cerfon. “Elimination of MHD Current Sheets by Modifications to the Plasma Wall in a Fixed Boundary Model.” Simons Workshop on Singularities in Fluids and Plasmas, Simons Collaboration on Hidden Symmetries and Fusion Energy, Princeton Center for Theoretical Science, Princeton, NJ, March 18-20, 2020.
16. J. Klobusicky. “Boundary Coarsening in Two Dimensions.” University of Scranton, Scranton, PA, January 31, 2020.
17. P. S. Kuo. “Zincblende nonlinear crystals for quantum information applications.” IEEE Photonics Conference, San Antonio, TX, October 2, 2019.
18. P. S. Kuo. “Entangled Photons for Quantum Information Applications.” Institute for Optical Science seminar, Ohio State University, Columbus, OH, November 8, 2019.
19. P. S. Kuo. “What's Next in Integrated Photonics – Hot Topics at CLEO: 2020.” Conference on Lasers and Electro-Optics, Online, May 11, 2020.
20. L. Ma. “From Quantum Cryptography to the Quantum Internet – an Overview of Quantum Communications and Networks.” Joint Seminar of iQUIST and iOptics, University of Illinois, Champaign-Urbana, IL, October 9, 2019.
21. V. Marbukh. “Towards Scalable Optimization of Large-Scale IoT Access for Mobile Users with Diverse Requirements.” The IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’2020), Online, August 31 - September 3, 2020.
22. D. Porter. “Tcl Core Team Year in Review.” 26th Annual Tcl/Tk Conference, Houston, TX, November 7, 2019.
23. B. Saunders. “Mathematics Behind the 3D Visualizations of the NIST Digital Library of Mathematical Functions.” George Mason University, October 25, 2019.
24. B. Saunders. “Mathematics, Mesh Generation, and 3D Graphics on the Web, and Finding a Career at NIST.” MAA MD-DC-VA Fall Section Meeting, Norfolk State University, November 9, 2019.
25. B. Saunders. “Celebrating the Career of Dr. Fern Hunt.” MAA Session on The EDGE (Enhancing Diversity in Graduate Education) Program: Pure and Applied Talks by Women Math Warriors, 2020 Joint Mathematics Meetings, Denver, Colorado, January 16, 2020.
26. B. Saunders, S. Brooks, R. Buckmire, and R. Vincent-Finley. “Validated Numerical Computation of Mathematical Functions.” African Diaspora Joint Mathematics Workshop (ADJOINT) 2020, Online, June 26, 2020.
27. B. Saunders. “From Abramowitz and Stegun to the NIST Digital Library of Mathematical Functions and Beyond.” History of Mathematics Class, Prof. Hortensia Soto, Colorado State University, Online, November 10, 2020.
28. K. Sayrafian. “IoT for Health: The Communication Infrastructure for a Pervasive Healthcare Environment.” IRACON Workshop, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, January 28, 2020.
29. S. Pamidighantam, B. Schneider, K. Bartschat, O. Zatsarinny, K. Hamilton, I. Bray, A. Scrinzi, F. Martin, J. Vasquez, J. Tennyson, J. Gorfinkiel, and R. Lucchese. “Atomic and Molecular Scattering Applications in an Apache Airavata Science Gateway.” PEARC ‘20: Practice and Experience in Advanced Research Computing, Online, July 29, 2020.

30. O. Slattery. "Quantum Repeaters: What They Are and Why They Matter." Quantum Network Grand Challenge Seminar Series, NIST, Online, May 22, 2020.
31. J. Zwolak. "Auto-Tuning of Quantum Dot Devices: Two Dots and Beyond." The Hume Center Tech Talk, Virginia Tech, Arlington, VA, February 10, 2020.
32. J. Zwolak. "Machine Learning Supported Auto-Tuning of Quantum Dot Devices: Two Dots and Beyond." Special Seminar at the Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, December 2, 2019.
33. J. Zwolak. "Machine Learning for Automated Formation of Quantum Dot Arrays." The Northwest Workshop on Quantum Transduction, Seattle, WA, November 15, 2019.

Conference Presentations

1. V. V. Albert. "Robust Encoding of a Qubit in a Molecule." Physical Review Journal Club, American Physical Society, Online, November 23, 2020.
2. V. V. Albert. "Robust Encoding of a Qubit in a Molecule." NSF Physics Frontier Center, Caltech, Online, November 18, 2020.
3. A. Avagyan. "State Tomography with Photon Counting After a Beam Splitter." Southwest Quantum Information and Technology Workshop, Eugene, OR, February 8-10, 2020.
4. B. Cloteaux. "Fast Graphic Approximation of Very Large Integer Sequences." International School and Conference on Network Science (NetSci) 2020, Online, September 22, 2020.
5. M. Coudron and S. Menda. "Computations with Greater Quantum Depth Are Strictly More Powerful (Relative to an Oracle)." 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2020), Online, June 2020.
6. M. Donahue and D. Porter. "High Order Methods for Computing the Demagnetization Tensor for Periodic Boundaries." 65th Annual Conference on Magnetism and Magnetic Materials, Online, November 6, 2020.
7. J. Klobusicky. "Particle System Methods for Networks with Two Types of Coarsening." World Congress in Computational Mechanics (WCCM) and European Community on Computational Methods in Applied Sciences (ECCOMAS) Congress 2020 Conference, Online, November 21, 2020.
8. A. Kwiatkowski. "Constraints on Continuous Variable Entanglement Swapping in the Presence of Photon Loss." American Physical Society March Meeting, Denver, Colorado, March 2-6, 2020.
9. S. Ressler. "Too Hot to Handle: Web3D for Fire Data Exploration." 12th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia, Brisbane, Australia, October 24, 2019.
10. M. Ruiz, J. Sims, and B. Padhy. "Exponentially Correlated Hylleraas-Configuration-Interaction Nonrelativistic Energy of the singlet S Ground State of the Helium Atom." Warsaw Electronic Structure Virtual Conference, Online, September 1-4, 2020.
11. K. Sayrafian. "A Machine Learning Approach to the Estimation of Near-Optimal Electrostatic Force in Micro Energy-Harvesters." IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE'19), Ottawa, Canada, October 16-18, 2019.
12. K. Sayrafian. "Utility Proportional Resource Allocation for Users with Diverse SLAs in Virtualized Radio Access Networks." IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom'2020), Online, May 26-29, 2020.
13. J. Zwolak, S. Kalantre, T. McJunkin, E. MacQuarrie, D. Savage, M. Lagally, M. Eriksson, and J. Taylor. "Ray-Based State Learning of Double Quantum Dot Systems." Silicon Quantum Electronics Workshop 2019, San Sebastian, Spain, October 14-16, 2019.

Poster Presentations

1. S. Bhushan. "Terahertz Electromagnetically Induced Transparency in Cesium Atoms." Conference on Frontier in Optics and Laser Science (FIO+LS), Online, September 14-17, 2020.
2. G. Doğan. "Scikit-Shape: Python Toolbox for Shape Analysis and Segmentation." SIAM Conference on Imaging Science, Online, July 6-17, 2020.
3. M. Donahue and D. Porter. "Quantitative Evaluation and Reduction of Error in Computation of the Demagnetization Tensor." 64th Annual Conference on Magnetism and Magnetic Materials, Las Vegas, NV, November 7, 2019.

4. S. Geller. "Reducing Ion Measurement Errors with Bayesian Techniques." Southwest Quantum Information and Technology Workshop, Eugene, Oregon, February 8-10, 2020.
5. T. Gerrits. "Calibration of Free-Space and Fiber-Coupled Single Photon Detectors." Conference on Quantum 2.0, Online, September 14-17, 2020.
6. W. Griffin, K. Sayrafian, and K. Krhac. "Sample Location Selection for Statistical Pathloss Modeling of Wireless Capsule Endoscopy." IEEE Visualization Conference (IEEE VIS'2019), Vancouver, Canada, October 20-25, 2019.
7. K. Krhac, K. Sayrafian, and D. Simunic. "A 3D Immersive Platform to Study Wireless Communication and Tracking of Ingestible Electronics." IEEE Engineering in Medicine and Biology (EMB) Special Topic Conference on Healthcare Innovations and Point-of-Care Technologies, National Institute of Health (NIH), Bethesda, Maryland, November 20-22, 2019.
8. K. Krhac, K. Sayrafian, and D. Simunic. "A Preliminary Study of the Human Body Communication Channel." IEEE Engineering in Medicine and Biology (EMB) Special Topic Conference on Healthcare Innovations and Point-of-Care Technologies, National Institute of Health (NIH), Bethesda, Maryland, November 20-22, 2019.
9. P. S. Kuo. "Effect of Amplitude Mismatch on Entanglement Visibility in Photon-Pair Sources." OSA Advanced Photonics Congress, Online, July 13-16, 2020.
10. A. Kwiatkwoski. "Constraints on Photon Loss for Continuous-Variable Entanglement Swapping Using Gaussian Resources." Southwest Quantum Information and Technology Workshop, Eugene, Oregon, February 8-10, 2020.
11. L. Ma, X. Tang and O. Slattery. "Quantum Memory in Anti-Relaxation Coated Gas Cell." Single Photon Workshop, Milan, Italy, October 21-25, 2019.
12. N. Martys, W. George, R. Murphy, and K. Weigandt. "Pipe Flow of Sphere Suspensions Having a Power-law-dependent Fluid Matrix." Gordon Research Conference on Cutting-Edge Developments and Characterization of Cement-Based Materials, Ventura, California, February 23-28, 2020.
13. A. Moorthy, A. Kearsley, W. Mallard, W. Wallace, and S. Stein. "Inferring the Molecular Mass of an Analyte from its Electron Ionization Mass Spectrum." American Society of Mass Spectrometry Annual Meeting, Online, June 3, 2020.
14. J. Zwolak, T. McJunkin, S. Kalantre, J. Dodson, E. MacQuarrie, D. Savage, M. Lagally, S. Coppersmith, M. Eriksson, and J. Taylor. "Machine Learning for Automated Formation of Quantum Dot Arrays." QuICS Five-Year Anniversary Symposium, College Park, MD, January 22, 2020.

Multimedia

1. S. Ressler. DLMF in VR. Interactive WebVR application. URL: https://math.nist.gov/~SRessler/dlmfvr.html
2. S. Ressler. NIST VR Scanning Tunneling Microscopy Head. Interactive WebVR application. URL: https://math.nist.gov/~SRessler/stm_scene/
3. S. Ressler. Virtual Tours of the NIST Library, Shops, and Net Zero House. URL: https://math.nist.gov/~SRessler/jab15/HomePage/
4. S. Ressler. Crown Fire. Immersive video. URL: https://math.nist.gov/~SRessler/aframe/CAVEvids/crown-Fire.html
5. S. Ressler. Photogrammetry Testing: Model of Great Synagogue of Florence. March 30, 2020. URL: https://sketchfab.com/3d-models/model-of-the-great-synagogue-of-florence-0f957129651149dfa3b7e7612a830865
6. S. Ressler. Test of the Scanning Tunneling Microscope Data in New Sketchfab Environment with Annotation Viewpoints. November 12, 2020. URL: https://sketchfab.com/3d-models/stm-head-all-947bbe0d4eb0460dbd6c9efed63ee27d

NIST News Releases

The following news items released by the NIST Public Affairs Office describe work in which ACMD staff members have participated.

1. Counting Photons Is Now Routine Enough to Need Standards. December 20, 2019.
2. To Tune Up Your Quantum Computer, Better Call an AI Mechanic. March 31, 2020.
3. NIST's Post-Quantum Cryptography Program Enters 'Selection Round'. July 22, 2020.
4. NIST Innovation Could Improve Detection of COVID-19 Infections. October 5, 2020.
5. Thermal MagIC: New NIST Project to Build Nano-Thermometers Could Revolutionize Temperature Imaging. October 9, 2020.
6. Error-Prone Quantum Bits Could Correct Themselves, NIST Physicist Show. December 8, 2020.

Web Services

Note: ACMD provides a variety of information and services on its website. Below is a list of major services provided that are currently under active maintenance.

1. *Digital Library of Mathematical Functions:* a repository of information on the special functions of applied mathematics.
2. *Digital Repository of Mathematical Formulae*: a repository of information on special function and orthogonal polynomial formulae.
3. *DLMF Standard Reference Tables on Demand*: an online software testing service providing tables of values for special functions, with guaranteed accuracy to high precision.
4. *muMAG*: a collection of micromagnetic reference problems and submitted solutions.

Software Released

Note: ACMD distributes a large number of software packages that have been developed in the course of its work. Listed below are particular packages which have seen new releases during the reporting period.

1. *ACTS:* Advanced Combinatorial Testing System. Version 3.2. R. Kacker.
2. *Elastic Registration and Shape Distance*. Version 1 (11/30/2020). J. Bernal.
3. *FMM3D*: A set of libraries to compute N-body interactions governed by the Laplace and Helmholtz equations, to a specified precision, in three dimensions, on a multi-core shared-memory machine, Version 1.0.0. Z. Gimbutas, L. Greengard, L. Lu, M. Rachh, and V. Rokhlin.
4. *Itcl*: C++ inspired object-oriented commands for Tcl. Versions 4.2.0 (11/21/19). D. G. Porter.
5. *LaTeXML*: A LaTeX to XML converter. Continuous access from svn/git repository. Version 0.8.5 (11/17/2020). B. R. Miller.
6. *Model-Based Optical Metrology in R*. Version 1 (10/30/2019). M.-A. Henn.
7. *A Library to Enable the Modeling of Optical Imaging of Finite Multi-Line Arrays*. Version 1 (10/30/2019). M.-A. Henn.
8. *MPI_GEVP*: Software for Solving Large-scale Generalized Eigenvalue Problems on Distributed Computers, Version 1 (9/9/2020). J. S. Sims.
9. *MWrap*: A MEX interface generation system in the spirit of SWIG or matwrap. Version 1.0 (8/5/20). D. Bindel, Z. Gimbutas, A. Barnett, L. Lu.
10. *OOF2:* Image-Based Analysis of Materials with Complex Microstructures, Version 2.1.19 (9/9/20). S. A. Langer, A. C. E. Reid, and G. Dogan.
11. *OOF3D*: Three-Dimensional Analysis of Materials with Complex Microstructures, Version 3.2.3 (4/30/20). S. A. Langer, A. C. E. Reid, and G. Dogan.
12. *OOMMF*: The Object Oriented MicroMagnetic Framework, Versions 1.2b4 (9/30/20). M. Donahue and D. Porter.
13. *scikit-shape*: Python package for shape and image analysis, Version 0.1 (7/26/20). G. Doğan.
14. *Sqlite3*: Bindings to the SQLite database engine for Tcl. Versions 3.30.1 (11/21/19). D. Porter.
15. *Tcl/Tk*: Extensible scripting language and GUI toolkit. Versions 8.6.10 (11/21/19), 8.7a3 and 9.0a1 (11/25/19). D. G. Porter.
16. *TDBC*: Database connection commands for Tcl. Version 1.1.1 (11/21/19). D. G. Porter.
17. *Thread*: Thread management commands for Tcl. Versions 2.8.5 (11/21/19), and 3.0a1 (11/25/19). D. G. Porter.
18. *VEMOS*: Visual Explorer for Metrics of Similarity, Version 1.1.3 (9/27/20), E. Fleisig and G. Doğan.

Data Released

1. A. Youssef and B. Miller. Math-dlmf-dataset, 2020. URL: https://github.com/abdouyoussef/math-dlmf-dataset

Conferences, Minisymposia, Lecture Series, Courses

ACMD Seminar Series

Stephen Langer served as Chair of the ACMD Seminar Series. There were 17 talks presented during this period; talks are listed chronologically.

1. Endel Iarve (University of Texas, Arlington). Regularized Finite Element Modeling Methodology for Predicting Discrete Damage-Based Strength and Durability of Laminated Composite Structures. October 8, 2019.
2. Marisabel Rodriguez (Arizona State University). Mathematical Models for Honeybee Population Dynamics. October 21, 2019.

3. Thomas Brown (George Mason University). Two Applications of PDE Constrained Optimization. November 12, 2019.
4. Daniel Nunez (LPI, Inc.). On Entropy Inequality and ordered Rate Constitutive Theories for Homogeneous, Isotropic, and Compressible Finite Deformation with Thermal Effects. December 3, 2019.
5. Hasan Eruslu (University of Delaware). FEM-CQ Approximation of Viscoelastic Waves. December 11, 2019.
6. Sharon Gourdji and Subhomoy Ghosh (NIST Greenhouse Gas and Climate Science Measurements Program). Tailoring Atmospheric Inversions for Estimating Greenhouse Gas Emissions in Urban Areas. January 13, 2020.
7. Robert DeJaco (University of Minnesota). Modeling Chemical Separations in All-Silica Zeolite Crystals and Spiral-Wound Membrane Modules. February 21, 2020.
8. Luis Dorfmann (Tufts University). Electromechanics of Soft Materials. February 25, 2020.
9. Zachary Grey (ACMD). Subspace-based Parameter Dimension Reduction and manifold Extensions. March 3, 2020.
10. Maryam Yashtini (Georgetown University). Efficient Variational Methods for Parallel MRI Reconstruction using Sensitivity Encoding. March 10, 2020.
11. Lucas Brady (ACMD). QA vs. QAO vs. QAOA: Optimizing Analog Quantum Computations. April 29, 2020.
12. Zachary Levine (NIST/PML). Tomography and Optimization using Monte Carlo. May 14, 2020.
13. Biswadip Dey (Siemens Corporate Technology). Physics-informed Machine Learning to Infer Dynamics from Data. May 21, 2020.
14. Bolong Zhang (Florida State University). Pass-efficient Randomized LU Algorithms for Computing Low-rank Matrix Approximations. May 26, 2020.
15. Jeffrey Hokanson (University of Colorado). Using the Lipschitz Matrix for Dimension Reduction. June 16, 2020.
16. Kamran Sayrafian (ACMD). A Study of the RF Propagation Channel for Wireless Capsule Endoscopy. August 11, 2020.
17. Mark-Alexander Henn (ACMD). Challenges in the Application of Data-driven Methods in Optical Metrology. August 11, 2020.

Shortcourses

1. M. Donahue. "Micromagnetics and OOMMF." Online Spintronics Seminar, May 21, May 26, June 2, and June 9, 2020.

Conference Organization

Leadership

1. M. Mascagni. Co-Organizer, Workshop on Computational Reproducibility at the Exascale, SC19, Denver, CO, November 17, 2019.
2. B. Miller. Co-Chair, Conference on Intelligent Computer Mathematics (CICM), Online, July 26-31, 2020.
3. S. Ressler. Demo Chair, Virtual Reality Continuum and its Applications in Industry 2019, Brisbane Australia, November 14-16. 2019.
4. K. Sayrafian. Panel Co-Chair, IEEE Conference on Standards for Communications and Networking (CSCN'19), Granada, Spain, October 29-31, 2019.
5. K. Sayrafian. Co-Chair, Vertical Applications and Internet of Things Track, IEEE European Conference on Networks and Communications (EuCNC'21) 2021, Porto, Portugal, June 8-11, 2021.
6. K. Sayrafian. Lead Organizer, Third International Workshop on IoT Enabling Technologies in Healthcare (IoT-Health'20), Online, June 11, 2020.
7. K. Sayrafian. Lead Organizer, Fourth International Workshop on IoT Enabling Technologies in Healthcare (IoT-Health'21), Montreal, Canada, June 14, 2021.
8. B. Schneider. Local Organizer, Workshop on A Science Gateway for Atomic and Molecular Physics, NIST, Gaithersburg, MD, December 11-13, 2019.

Committee Membership

1. M. Coudron. Member, Program Committee, 23rd Annual Conference on Quantum Information Processing (QIP), Shenzhen, China, January 6-10, 2020.
2. M. Coudron. Member, Program Committee, 16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC), Online, July 5-8, 2021.
3. T. Gerrits. Member, Technical Program Committee, Quantum Technologies, SPIE Photonics Europe 2020, Online, April 6-10, 2020.

4. T. Gerrits. Member, Topical Subcommittee, Quantum Information and Communications, Conference on Lasers and Electro-Optics (CLEO) 2020, Online, May 10-15, 2020.
5. T. Gerrits. Member, Topical Subcommittee, Quantum Information and Communications, Conference on Lasers and Electro-Optics (CLEO) 2021, San Jose, CA, May 9-14, 2021.
6. R. Kacker. Member, Steering Committee, 9th International Workshop on Combinatorial Testing (IWCT 2020), 13th IEEE International Conference on Software Testing, Verification, and Validation (ICST 2020), Online, October 24-28, 2020.
7. P. S. Kuo. Member, Program Committee. 2020 SPIE Photonics West: Advanced Optical Techniques for Quantum Information, Sensing, and Metrology, San Francisco, CA February 4-5, 2020.
8. P. S. Kuo. Member, Program Sub-committee. 2020 Conference on Lasers and Electro-Optics: Quantum Photonics, Online, May 11-15, 2020.
9. P. S. Kuo. Member, Program Committee. 2020 OSA Nonlinear Photonics, Online, July 13-16, 2020.
10. R. La. Member, Technical Program Committee, 10th International Conference on Complex Networks and Their Applications (Complex Networks 2020), Online, December 1-3, 2020.
11. R. La. Member, Technical Program Committee, IEEE International Conference on Computer Communications (INFOCOM 2010), Online, May 10-13, 2021.
12. K. Sayrafian. Member, Technical Program Committee, 15th EAI International Conference on Body Area Networks (BODYNETS 2020), Online, October 21-22, 2020.
13. K. Sayrafian. Member, International Advisory Board, International Symposium on Medical Information and Communication Technology (ISMICT).
14. B. Schneider. Member, Planning Committee, NITRD Workshop: Software in the Era of Extreme Heterogeneity, Online, September 22-24, 2020.
15. O. Slattery. Program Committee, Quantum Communications and Imaging XI, SPIE Optics and Photonics, Online, August 24-28, 2020.

Session Organization

1. H. Cohl, Co-Organizer, Special Session on the Legacy of Dick Askey, 2021 Joint Mathematics Meetings, Online, January 6-9, 2021.
2. G. Doğan. Co-Organizer, Minisymposium 53, Advances in Variational Models and PDEs for Images. SIAM Conference on Imaging Science, Online, July 6-17, 2020.
3. T. Gerrits. Co-Organizer, Quantum Biophotonics Symposium, Conference on Lasers and Electro-Optics (CLEO) 2020, Online, May 10-15, 2020.
4. S. Ressler. Organizer, Immersive Analytics Birds-of-a-Feather, SIGGRAPH Asia 2019, Brisbane, Australia, November 17-20, 2019.
5. K. Sayrafian. Session Co-organizer, Union Radio-Scientifique Internationale (URSI) General Assembly and Scientific Symposium (URSI GASS 2021), Rome, Italy, August 28 - September 4, 2021.

Other Professional Activities

Internal

1. R. Boisvert. NIST Quantum Networking Grand Challenge Planning Team.
2. R. Boisvert. Chair, NIST Research Computing Infrastructure Team.
3. B. Cloteaux. Member, Washington Editorial Review Board.
4. M. Coudron. Co-Organizer, QuICS Weekly Research Seminar.
5. S. Glancy. Member, Boulder Summer Undergraduate Research Fellowship Committee.
6. S. Glancy. Member, ITL Diversity Committee.
7. A. Kearsley. Chair, ITL Awards Committee.
8. P. Kuo. NIST Quantum Networking Grand Challenge Planning Team.
9. P. Kuo. Division Representative, ITL Space Task Force.
10. S. Langer. Chair, ACMD Seminar Series.
11. L. Ma. Organizer, NIST Quantum Repeater Journal Club.
12. M. McClain. Division Property Officer.
13. L. Orr. Member, ITL Diversity Committee.
14. L. Orr. Division Property Officer.
15. D. Porter. Member ITL Awards Committee
16. S. Ressler. Division Safety Representative.
17. B. Schneider. Member, NIST Research Computing Infrastructure Team.

18. O. Slattery. Laser Safety Representative, ITL Safety Committee.
19. O. Slattery. Division Representative, ITL Space Task Force.
20. O. Slattery. ITL Representative, NIST Laser Safety Committee.
21. J. Zwolak. At-Large Member, Steering Committee, NIST AI Community of Interest.

External

Editorial

1. R. Boisvert. Associate Editor, *ACM Transactions on Mathematical Software*.
2. H. Cohl. Member, Editorial Board, *The Ramanujan Journal*.
3. H. Cohl. Co-Editor, OP-SF NET Newsletter, SIAM Activity Group on Orthogonal Polynomials and Special Functions.
4. H. Cohl. Guest-Editor, *Symmetry*, Special Issue on Symmetry in Special Functions and Orthogonal Polynomials.
5. H. Cohl. Co-Editor, *Liber Amicorum*, a Friendship Book for Dick Askey.
6. H. Cohl. Co-Editor, Richard A. Askey Memorial Volume, *The Ramanujan Journal*.
7. T. Gerrits. Associate Editor, *Optics Express*.
8. Z. Gimbutas. Member, Editorial Board, *Advances in Computational Mathematics*.
9. S. Glancy. Associate Editor, *Quantum Information Processing*.
10. R. La. Associate Editor, *IEEE/ACM Transactions on Networking.*
11. D. G. Porter. Member, Editorial Board, *Journal of Research of NIST*
12. F. Potra. Regional Editor for the Americas, *Optimization Methods and Software.*
13. F. Potra. Associate Editor, *Journal of Optimization Theory and Applications*.
14. F. Potra. Associate Editor, *Numerical Functional Analysis and Optimization*.
15. F. Potra. Associate Editor, *Optimization and Engineering*.
16. B. Saunders. Associate Editor, *MAA Mathematics Magazine*.
17. B. Saunders. OP-SF Talk Listserv Moderator, SIAM Activity Group on Orthogonal Polynomials and Special Functions.
18. K. Sayrafian. Associate Editor, *International Journal on Wireless Information Networks*.
19. B. Schneider, Associate Editor-in-Chief, *IEEE Computing in Science & Engineering*.

Boards and Committees

1. R. F. Boisvert. Member, ACM Digital Library Committee.
2. R. Boisvert. Member, Working Group 2.5 (Numerical Software), International Federation for Information Processing.
3. B. Cloteaux. Member, Advisory Board, Department of Computer Science, New Mexico State University.
4. A. Dienstfrey. Member, Working Group 2.5 (Numerical Software), International Federation for Information Processing.
5. T. Gerrits. Member, Scientific Board, Single-photon Sources as New Quantum Standards (SIQUST) Project, European Union.
6. S. Glancy. Member, IEEE Working Group on Metrics and Benchmarks for Quantum Computing Devices and Systems.
7. P. S. Kuo. Member, Executive Committee, American Physical Society, Mid-Atlantic Section,
8. B. Miller. Member, W3C Math Working Group.
9. P. Patrone. Member, Applied Math and Scientific Computing Graduate Committee, University of Maryland.
10. D. Porter. Member, Tcl Core Team.
11. S. Ressler. Member, W3C Immersive Web Working Group.
12. S. Ressler. NIST Advisory Committee member W3C.
13. S. Ressler. NIST Representative. Khronos Group.
14. B. Saunders. Member, Board of Trustees, Society for Industrial and Applied Mathematics (SIAM).
15. B. Saunders. MD-DC-VA Section Representative, Congress of Mathematical Association of America (MAA).
16. B. Saunders. Webmaster, SIAM Activity Group on Orthogonal Polynomials and Special Functions.

17. K. Sayrafian. Co-Chair, IoT-Health Technical Group, COST CA115104 Inclusive Radio Communication Networks for 5G and Beyond (IRACON).
18. B. Schneider. Co-Chair, OSTP Fast Track Action Committee on Strategic Computing.
19. B. Schneider. NIST Representative, NITRD High End Computing Working Group.

Adjunct Academic Appointments

1. V. V. Albert. Adjunct Assistant Professor, Department of Physics, University of Maryland, College Park, MD.
2. V. V. Albert. Adjunct Assistant Professor, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD.
3. M. Coudron. Adjunct Assistant Professor, Department of Computer Science, University of Maryland, College Park, MD.
4. M. Coudron. Adjunct Assistant Professor, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD.
5. S. Glancy. Lecturer, Department of Physics, Colorado University, Boulder, CO.
6. E. Knill. Lecturer, Department of Physics, Colorado University, Boulder, CO.
7. P. Kuo. Adjunct Associate Professor, Department of Physics, University of Maryland, College Park, MD.
8. Y.-K. Liu. Adjunct Associate Professor, Department of Computer Science, University of Maryland, College Park, MD.
9. Y.-K. Liu. Adjunct Associate Professor, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD.
10. K. Sayrafian. Affiliate Associate Professor, Concordia University, Montreal, Canada.
11. X. Tang. Adjunct Professor, University of Limerick, Ireland.

Thesis Direction

1. G. Doğan. Member, Ph.D. Thesis Committee, University of Delaware: H. H. Eruslu.
2. M. Donahue. Member, Ph.D. Thesis Committee, University of Colorado, Boulder: Mingyu Hu.
3. Y.-K. Liu. Member, Ph.D. Thesis Committee, University of Maryland, College Park: T. Li, Quantum Algorithms for Machine Learning and Optimization, March 2020.
4. Y.-K. Liu. Member, Ph.D. Thesis Committee, University of Maryland, College Park: A. Seif, Control and Characterization of Open Quantum Systems, October 2020.
5. Y.-K. Liu. Co-Advisor, University of Maryland, College Park: K. Huang.
6. F. Potra. Advisor, University of Maryland Baltimore County: J. Praniewicz.
7. K. Sayrafian. Co-Advisor, University of Zagreb, Zagreb, Croatia: K. Krhac.
8. K. Sayrafian. Co-Advisor, Concordia University, Montreal, Canada: M. Roudneshin.
9. K. Sayrafian. Examiner/Opponent, Ph.D. Thesis Committee, University of Oulu: M. Särestöniemi.

Community Outreach

1. R. Evans. Question and Answer Panelist. Applied Mathematics and Statistics, and Scientific Computing Program Visiting Day. University of Maryland, College Park, MD, March 9, 2020.
2. B. Saunders. SIAM Visiting Lecturer.
3. B. Saunders. Research Leader, Validated Numerical Computations of Mathematical Functions, African Diaspora Joint Mathematics Workshop (ADJOINT) 2020, Mathematical Sciences Research Institute, Online, June 15-26, 2020.
4. B. Saunders. Advisor, Virginia Standards of Learning Tutoring Program, Northern Virginia Chapter (NoVAC), Delta Sigma Theta Sorority, Inc. and the Dunbar Alexandria-Olympic Branch of the Boys and Girls Clubs of Greater Washington, January – May 2020.

Awards and Recognition

External

1. R. Boisvert. Fellow, Association for Computing Machinery, December 2019.
2. M. J. Donahue. Excellence in Research in Applied Mathematics, Washington Academy of Sciences, September 10, 2020.
3. M. J. Donahue. Fellow, Washington Academy of Sciences, September 10, 2020.

4. S. Glancy. Outstanding Referee. American Physical Society, December 2020.
5. F. Hunt. Fellow, Association for Women in Mathematics, November 2019.
6. D. Mitra. Fellow, American Association for the Advancement of Science, December 2020.
7. K. Sayrafian. Certificate of Appreciation, IEEE Conference on Standards for Communications and Networking (CSCN'19), Granada, Spain, October 29, 2019.

Internal

8. B. Cloteaux and V. Marbukh. ITL Outstanding Conference Paper Award, July 21, 2020.
9. W. George and S. Satterfield. NIST Judson C. French Award (Group award), 2020.
10. S. Glancy and E. Knill. DOC Gold Medal (Group award), 2020.
11. A. Kearsley and P. Patrone. NIST MML Collaboration and Teamwork Award (Group award), 2020.
12. D. Lozier. Inductee, NIST Portrait Gallery of Distinguished Scientists, Engineers and Administrators, October 2020.
13. G. McFadden. ITL Outstanding Contribution to Enhance Diversity Award, 2020.
14. P. Patrone. 2020 Outstanding Contribution to ITL Award, July 14, 2020.
15. B. Saunders. NIST Diversity, Inclusion and EEO Award, 2020.
16. B. Schneider. DOC Bronze Medal (Group award), 2020.

Funding Received

During FY 2020 ACMD's yearly allocation of base funding from the NIST Information Technology Laboratory was supplemented with funding from a variety of internal and external competitions. Such funding represented 12 % of the Division's FY 2020 budget.

Note: For multi-year awards and joint awards, only projects with new funding received by ACMD during FY 2020 are listed. For joint projects, names of ACMD participants are underlined.

External

1. M. Donahue. NIST Support for DARPA Magnetic Miniaturized and Monolithically Integrated Component (M3IC) Program, DARPA. (Joint with PML)

Internal

1. B. Cloteaux, V. Marbukh, and D. Mitra. Towards Actionable Cybersecurity Risk Metrics in a Hyperconnected World. ITL Building the Future Program.
2. C. L. Dennis, T. P. Moffat, A. J. Biacchi, A. R. Hight Walker, S. I Woods, W. L. Tew, and M. J. Donahue. Thermal MagIC: An SI-Traceable Method for 3D Thermal Magnetic Imaging and Control. NIST Innovations in Measurement Science. (Joint with PML, MML)
3. A. Dienstfrey and A. Mink. Sparse by Spike: Spiking Neural Networks for Sparse Recovery. ITL Building the Future Program.
4. R. Evans, A. Balijepalli, C. Schanzle, and A. Kearsley. Modeling, Uncertainty Quantification, and Calibration for BioFETs. ITL Building the Future Program.
5. R. Fitzgerald, D. Bergeron, S. Nour, D. Schmidt, D. Swetz, G. Shaw, B. Alpert, and M. Verkouteren. True Becquerel: A New Paradigm for 21st Century Radioactivity Measurements (Joint with PML)
6. D. Leibfried, A. Wilson, E. Knill, and S. Glancy. A Practical Quantum Repeater Unit. NIST Quantum Networking Grand Challenge. (Joint with PML)
7. K. Lehnert, K. Silverman, D. Moody, J. Teufel, R. Mirin, S.-W. Nam, E. Knill, P. Hale, and T. Dennis. Establishing the Science of Networks for Superconducting Quantum Computers. NIST Innovations in Measurement Science Program. (Joint with CTL, PML)

8. P. Patrone, A. Kearsley, G. McFadden, G. Cooksey, S. Sarkar, and L. Wang. NIST-in-a-drop: Revolutionizing Measurements of Single-cell Kinetics. NIST Innovations in Measurement Science. (Joint with PML, MML)
9. J. Porto, A. Gorshkov, A. Migdall, W. Phillips, S. Glancy, E. Knill, and K. Srinivasan. Metrology with Interacting Photons. NIST Innovations in Measurement Science Program. (Joint with PML)
10. K. Sayrafian and W. Griffin. A Statistical Pathloss Model for Ingestible Wireless Devices Using Electric Field Communication. ITL Building the Future Program.
11. X. Tang, L. Ma, and O. Slattery. A New Optical Fiber Device for Coupling with Micro Resonators to Generate Single Photon Pairs. ITL Building the Future Program.
12. A. Youssef and B. Miller. Deep Learning Neural Networks for Math and Science Knowledge Processing. ITL Building the Future Program.

Grants Awarded

ACMD awards a small amount of funding through the NIST Measurement Science Grants Program for projects that make direct contributions to its research programs. During FY 2020 the following cooperative agreements were active.

1. Prometheus Computing LLC: *Security, Resiliency and Dynamics of Interdependent Self-Organizing Networks*. PI: Assane Gueye.
2. Theiss Research: *Exploiting Alternate Computing Technologies*. PI: Alan Mink.
3. University of Edinburgh: *Rigorous and Presentable Asymptotics for Special Functions and Orthogonal Polynomials*. PI: Adri Olde-Daalhuis.
4. University of Maryland: *In-situ Visualization for Immersive Environments*. PI: Amitabh Varshney.
5. University of Maryland: *Joint Center for Quantum Information and Computer Science (QuICS)*. PI: Andrew Childs.
6. University of Texas at Arlington: *SENTINEL: Security Interaction Testing for IoT Systems and Blockchains*. PI: Yu Lei.

External Contacts

ACMD staff members interact with a wide variety of organizations in the course of their work. Examples of these follow.

Industrial Labs

BAE Systems
Center for Integration of Medicine & Innovative Tech.
Computational Physics, Inc.
Fluidigm
Fraunhoffer IGD (Germany)
Gener8
HC-Photonics Corp.
Honeywell
Intel Corporation
Intelligent Automation Inc.
Kitware
Microsoft Research
NTT Corporation (Japan)
Xanadu Quantum Technologies, Inc.

Government/Non-profit Organizations

American Telemedicine Association
Argonne National Laboratory
Army Research Laboratory
Association for Computing Machinery
CSIRO (Australia)
Food and Drug Administration
Idaho National Laboratory
IEEE Computer Society
Institut de la Vision (France)
International Federation for Information Processing
Johns Hopkins University Applied Physics Laboratory
Lawrence Berkeley National Laboratory
Lawrence Livermore National Laboratory
Mathematical Sciences Research Institute
Nanohub.org
National Cancer Institute
National Institute of Biomedical Imaging and Bioengineering
National Institutes of Health
National Science Foundation
Quantum Economic Development Consortium
Sandia National Laboratories
Society for Industrial and Applied Mathematics
Theiss Research
US Department of Energy
US Geological Survey
US Holocaust Memorial Museum
US Nuclear Regulatory Commission

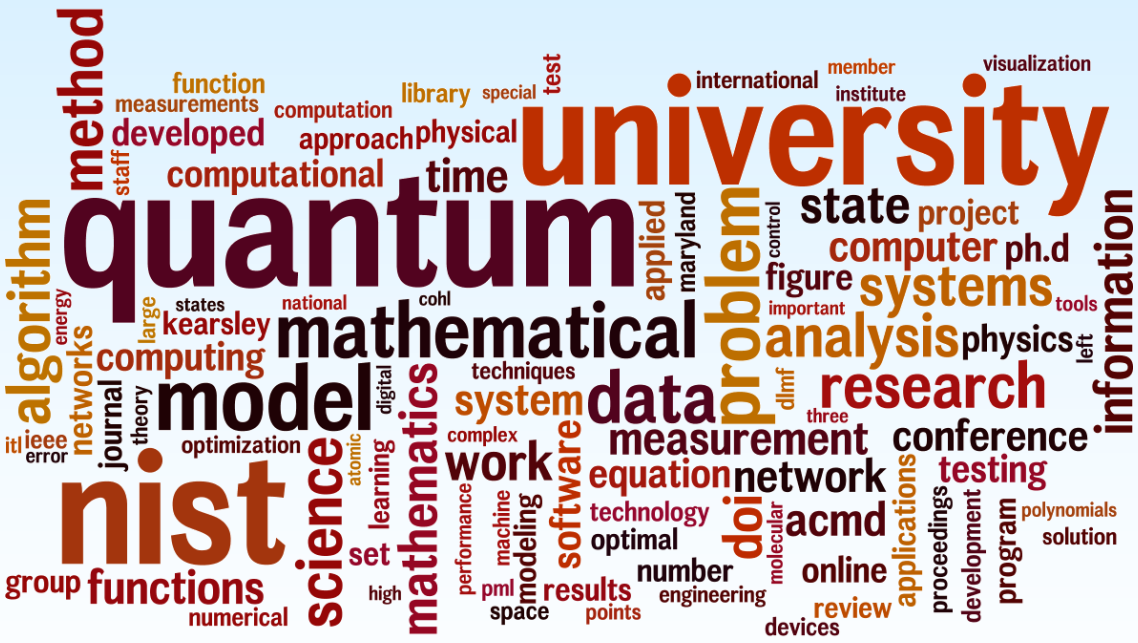
Universities

Aarhus University (Denmark)
American University

Arizona State University
Brown University
Caltech
Carnegie Mellon University
Clarkson University
Colorado School of Mines
Colorado State University
Columbia University
Comenius University (Slovakia)
Concordia University (Canada)
Coppin State University
Courant Institute of Mathematical Sciences
Curtin University (Australia)
Drexel University
Durham University
ETH Zurich (Switzerland)
European University Cyprus (Cyprus)
Federal University of Ceara (Brazil)
Flatiron Institute
Florida State University
Friedrich-Alexander University (Germany)
Freie Universität Berlin (Germany)
Gdansk University of Technology (Poland)
George Mason University
Georgetown University
George Washington University
Imperial College London (UK)
Indiana University
Jacobs University Bremen (Germany)
Kennesaw State University
Khallikote College (India)
Linnaeus University (Sweden)
Louisiana State University
Ludwig-Maximilians University (Germany)
Massachusetts Institute of Technology
Morgan State University
New York University
Norfolk State University
Occidental College
Ohio State University
Open University (UK)
Oxford University
Paderborn University (Germany)
Polytechnic University of Valencia (Spain)
Princeton University
Purdue University
Rollins College
Royal Institute of Technology (Sweden)
Stanford University
State University of New York
ShanghaiTech University (China)
Shippensburg University
Southern University and A&M College
SUNY Binghamton
Technical University of Berlin (Germany)
Texas A&M University
Universidad Autonoma de Madrid (Spain)
University of Akron
University of Antwerp (Belgium)
University of Bergamo (Italy)
University of Bologna (Italy)
University of California, Los Angeles
University of California, Santa Barbara
University of California, San Diego
University of Central Florida
University of Colorado, Boulder
University of Delaware
University of Edinburgh (UK)
University of Erlangen (Germany)
University of Freiburg (Germany)
University of Illinois, Urbana-Champaign
University of Limerick (Ireland)
University of Lisbon (Portugal)
University of Maryland, Baltimore County
University of Maryland, College Park
University of Melbourne (Australia)
University of Michigan
University of New Mexico
University of New Orleans
University of New South Wales (Australia)
University of Oregon
University of Oulu (Finland)
University of Saskatchewan (Canada)
University of Scranton
University of Southern California
University of Sydney (Australia)
University of Texas at Arlington
University of Texas at Austin
University of Toronto (Canada)
University of Waterloo (Canada)
University of Wisconsin
University of Wyoming
University of York (UK)
University of Zagreb (Croatia)
Virginia Commonwealth University
Virginia Polytechnic Institute
Worcester Polytechnic University
Yale University

Part V

Appendix



Staff

ACMD consists of full-time permanent Federal staff located at NIST laboratories in Gaithersburg, MD and Boulder, CO. This full-time staff is supplemented with a variety of special appointments. The following list reflects all non-student appointments held during any portion of the reporting period (October 2019 – December 2020). Students and interns are listed in Table 5 page 132.

* Denotes staff at NIST Boulder.
† Denotes part-time Federal staff.

Division Staff

Ronald Boisvert, *Chief*, Ph.D. (Computer Science), Purdue University, 1979
Catherine Graham, *Secretary*
Lochi Orr, *Administrative Assistant*, A.A. (Criminal Justice), Grantham University, 2009
†Alfred Carasso, Ph.D. (Mathematics), University of Wisconsin, 1968
Roldan Pozo, Ph.D. (Computer Science), University of Colorado at Boulder, 1991
Kamran Sayrafian-Pour, Ph.D. (Electrical and Computer Engineering), University of Maryland, 1999
Christopher Schanzle, B.S. (Computer Science), University of Maryland Baltimore County, 1989

Mathematical Analysis and Modeling Group

Timothy Burns, *Leader*, Ph.D. (Mathematics), University of New Mexico, 1977
*Daniel Flynn, *Administrative Assistant*, B.S. (Political Science), Iowa State University, 2016
*Bradley Alpert, Ph.D. (Computer Science), Yale University, 1990
*Andrew Dienstfrey, Ph.D. (Mathematics), New York University, 1998
Ryan Evans, Ph.D. (Applied Mathematics), University of Delaware, 2016
†Jeffrey Fong, Ph. D. (Applied Mechanics and Mathematics), Stanford University, 1966
*Zydrunas Gimbutas, Ph.D. (Applied Mathematics), Yale University, 1999
Raghu Kacker, Ph.D. (Statistics), Iowa State University, 1979
Anthony Kearsley, Ph.D. (Computational and Applied Mathematics), Rice University, 1996
Geoffrey McFadden, *NIST Fellow,* Ph.D. (Mathematics), New York University, 1979
Danielle Middlebrooks, Ph.D. (Applied Mathematics, Statistics and Scientific Computing), University of Maryland, 2020
Paul Patrone, Ph.D. (Physics), University of Maryland, 2013

Faculty Appointee (Name, Degree / Home Institution)
Daniel Anderson, Ph.D. / George Mason University
Michael Mascagni, Ph.D. / Florida State University
John Nolan, Ph.D. / American University
Florian Potra, Ph.D. / University of Maryland Baltimore County

NRC Postdoctoral Associates
Danielle Brager, Ph.D. (Mathematical Biology), Arizona State University, 2020
Robert DeJaco, Ph.D. (Chemical Engineering), University of Minnesota, 2020
*Zachary Grey, Ph.D. (Computational and Applied Math), Colorado School of Mines, 2019
Joseph Klobusicky, Ph.D. (Applied Mathematics), Brown University, 2014
Matthew Roberts, Ph.D. (Mathematical Sciences), Michigan Technological University, 2019

Guest Researchers (Name, Degree / Home Institution)
Sebastian Barillaro, Ph.D. / Industrial Technology National Institute, Argentina
Diego Andres Bouvier, M.S. / Technological Laboratory of Uruguay
Natesh Ganesh, Ph.D. / University of Colorado
Fern Hunt, Ph.D. / *NIST Scientist Emeritus*

Yu (Jeff) Lei, Ph.D. / University of Texas at Arlington
Dimitrios Simos, Ph.D. / SBA Research, Austria
Christoph Witzgall, Ph.D. / *NIST Scientist Emeritus*

Mathematical Software Group

Michael Donahue, *Leader* Ph.D. (Mathematics), Ohio State University, 1991
Javier Bernal, Ph.D. (Mathematics), Catholic University, 1980
Howard Cohl, Ph.D. (Mathematics), University of Auckland, 2010
Günay Doğan, Ph.D. (Applied Mathematics and Scientific Computing), University of Maryland, 2006
Stephen Langer, Ph.D. (Physics), Cornell University, 1989
Marjorie McClain, M.S. (Mathematics), University of Maryland College Park, 1984
Bruce Miller, Ph.D. (Physics), University of Texas at Austin, 1983
Donald Porter, D.Sc. (Electrical Engineering), Washington University, 1996
Bonita Saunders, Ph.D. (Mathematics), Old Dominion University, 1985
Barry Schneider, Ph.D. (Physics), University of Chicago, 1969

Faculty Appointees (Name, Degree / Home Institution)
Abdou Youssef, Ph.D. / George Washington University

Guest Researchers (Name, Degree / Home Institution)
Nicolas Douguet, Ph.D. / Kennesaw State University
Heman Gharibnejad, Ph.D. / Computational Physics Inc.
Deyan Ginev, M.Sc. / Chakra Consulting
Xiaoxu Guan, Ph.D. / Colorado School of Mines
Mark Alexander Henn, Ph.D. / University of Maryland
Justin Kauffman, Ph.D. / Virginia Polytechnic Institute and State University
Daniel Lozier, Ph.D. / NIST, Retired
Todd Martinez, Ph.D. / Stanford University
Adri Olde Daalhuis, Ph.D. / University of Edinburgh
Jeppe Olsen, Ph.D. / Aarhus University, Denmark
Chen Qu, Ph.D. / University of Maryland
Moritz Schubotz, Ph.D. / University of Karlsruhe, Germany

Computing and Communications Theory Group

Ronald Boisvert, *Acting Leader*, Ph.D. (Computer Science), Purdue University, 1979
Victor Albert, Ph.D. (Physics), Yale University, 2017
Brian Cloteaux, Ph.D. (Computer Science), New Mexico State University, 2007
Matthew Coudron, Ph.D. (Computer Science), Massachusetts Institute of Technology, 2017
Thomas Gerrits, Ph.D. (Physics), Radboud University Nijmegen, 2004
*Scott Glancy, Ph.D. (Physics), University of Notre Dame, 2003
*Emanuel Knill, *NIST Fellow,* Ph.D. (Mathematics), University of Colorado at Boulder, 1991
Paulina Kuo, Ph.D. (Physics), Stanford University, 2008
Yi-Kai Liu, Ph.D. (Computer Science), University of California, San Diego, 2007
Lijun Ma, Ph.D. (Precision Instruments and Machinery), Tsinghua University, 2001
Vladimir Marbukh, Ph.D. (Mathematics) Leningrad Polytechnic University, 1986
Oliver Slattery, *Project Leader*, Ph.D. (Physics), University of Limerick, 2015
†Xiao Tang, Ph.D. (Physics), Chinese Academy of Sciences, 1985

NRC Postdoctoral Associates
Lucas Brady, Ph.D. (Physics), University of California at Santa Barbara, 2018

Faculty Appointees (Name, Degree / Home Institution)
James Lawrence, Ph.D. / George Mason University

Richard La, Ph.D. / University of Maryland
Debasis Mitra, Ph.D. / Columbia University

Guest Researchers (Name, Degree / Home Institution)
Isabel Beichl, Ph.D. / NIST, Retired
Sumit Bhushan, Ph.D. / Indian Institute of Technology
*Bryan Eastin, Ph.D. / Northrup Grumman
Assane Gueye, Ph.D. / Prometheus Computing
Stephen Jordan, Ph.D. / Microsoft Research
Alan Mink, Ph.D. / Theiss Research
Anouar Rahmouni, Ph.D. / Moroccan Foundation for Advanced Sci., Innovation and Research
*Ezad Shojaee, Ph.D. / University of Colorado
Francis Sullivan, Ph.D. / IDA Center for Computing Sciences

High Performance Computing and Visualization Group

Judith Terrill, *Leader*, Ph.D. (Information Technology), George Mason University, 1998
William George, Ph.D. (Computer/Computational Science), Clemson University, 1995
Terence Griffin, B.S. (Mathematics), St. Mary's College of Maryland, 1987
Wesley Griffin, Ph.D. (Computer Science), University of Maryland Baltimore County, 2016
Sandy Ressler, M.F.A. (Visual Arts), Rutgers University, 1980
Steven Satterfield, M.S. (Computer Science), North Carolina State University, 1975
James Sims, Ph.D. (Chemical Physics), Indiana University, 1969
Justyna Zwolak, Ph.D. (Physics), Nicolaus Copernicus University, Poland, 2011

NRC Postdoctoral Associates
Joshua Ziegler, Ph.D. (Physics), University of Oregon, 2020

Guest Researchers (Name, Degree / Home Institution)
John Hagedorn, M.S. / Chakra Consulting
William Sherman, M.S. / Chakra Consulting

Glossary of Acronyms

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
ACM	Association for Computing Machinery
ACMD	NIST/ITL Applied and Computational Mathematics Division
ACTS	Advanced Combinatorial Testing System
ADJOINT	African Diaspora Joint Mathematics Workshop
ADLP	NIST Associate Director for Laboratory Programs
AI	artificial intelligence
AIAA	American Institute for Aeronautics and Astronomics
AM	analytical methods
AMP	atomic and molecular physics
AMS	American Mathematical Society
AMSC	Applied Mathematics and Statistics and Scientific Computing program at UMD
ANARI	Analytic Rendering Interface for Data Visualization
API	application programming interface
aPPLN	aperiodically poled lithium niobate
APS	American Physical Society
AR	augmented reality
AWM	Association for Women and Computing
arXiv	preprint archive housed at Cornell University (http://arxiv.org/)
Bio-FET	biological field effect transistor
BMP	Bateman Manuscript Project
BP	normal boiling point
BPP	class of decision problems solvable by a probabilistic Turing machine in polynomial time with an error probability bounded away from 1/3 for all instances
Bq	becquerel: absolute activity of radionuclide mixtures
BQP	class of decision problems solvable by a quantum computer in polynomial time
Caltech	California Institute of Technology
CAS	computer algebra system
CAVE	CAVE Automatic Virtual Environment
CCM	Combinatorial Coverage Measurement
CENAM	Center for Metrology of Mexico
CFPG	Coulomb force parametric generator
CI	configuration interaction
CICM	Conference on Intelligent Computer Mathematics
CLEO	Conference on Lasers and Electro-Optics
CMA	University of Antwerp Computational Mathematics Research Group
CPU	central processing unit
CFPG	Coulomb force parametric generator
CFSF	Continued Fractions for Special Functions
CMA	Computational Mathematics Research Group at the University of Antwerp
COVID	coronavirus disease
CPA	cryo-protective agent
CRADA	cooperative R&D agreement
CRD	cone-rod dystrophy
CSF	configuration state function
CSIRO	Australia's Commonwealth Scientific and Industrial Research Organization
CT	combinatorial testing
CTL	NIST Communications Technology Laboratory
CV	continuous variable
CVS	Concurrent Version System (for source code version control)
CY	calendar year

CyEVaR	entropic cybersecurity value at risk
CyVaR	cybersecurity value at risk
DARPA	Defense Advanced Research Projects Agency
DL	deep learning
DLMF	Digital Library of Mathematical Functions
DNA	deoxyribonucleic acid
DNN	deep neural network
DOC	Department of Commerce
DOI	digital object identifier
DSO	differential semblance optimization
DRMF	Digital Repository of Mathematical Formulae
EBSD	electron backscatter diffraction
eCF	Wolfram Computational Knowledge of Continued Fractions Project
EE	electrical engineering
EEO	equal employment opportunity
E-Hy-CI	exponentially correlated Hy-CI
EL	NIST Engineering Laboratory
EMD	earth mover distance
EPA	U.S. Environmental Protection Agency
EVar	entropic value at risk
FDA	Food and Drug Administration
FEDVR	finite element discrete variable
FEM	finite element method
FFT	fast Fourier transform
FGR	NIST Foreign Guest Researcher Program
FWHM	full width at half maximum
FY	fiscal year
GB	gigabyte
GC	gas chromatography
GDML	Global Digital Mathematical Library
GHG	greenhouse gas
GI	gastrointestinal
GNN	graph neural network
GPU	graphical processing unit
GRADES	Graph Data management Experiences and Systems (workshop)
GRN	gene regulatory network
GSVE	generalized singular value expansion
HEC	High End Computing
HEV	high end visualization
HMD	head-mounted display
HPCVG	ACMD High Performance Computing and Visualization Group
HTML	hypertext markup language
HVAC	heating, ventilation and air conditioning
HVACSIM+	software package and computing environment for simulating HVAC system
Hy-CI	Hylleraas-Configuration Interaction technique
IARPA	Intelligence Advanced Research Projects Agency
IBBR	UMD-NIST Institute for Bioscience and Biotechnology Research
ICST	International Conference of Software Testing
IDE	integro-differential equation
IEEE	Institute of Electronics and Electrical Engineers
IFIP	International Federation for Information Processing
IMKT	International Mathematical Knowledge Trust
IMS	NIST Innovations in Measurement Science program
IoT	Internet of things
IT	information technology
ITAMP	Harvard-Smithsonian Institute for Theoretical Atomic and Molecular Physics
ITL	NIST Information Technology Laboratory

IVE	immersive visualization environment
IWCT	International Workshop on Combinatorial Testing
KLS	Koekoek, Lesky and Swarttouw
KNN	K-nearest neighbors' algorithm
KPI	key performance indicator
KRI	Kováts retention index
KSU	Kansas State University
LAA	license-assisted access
LaTeX	a math-oriented text processing system
LaTeXML	a LaTeX to Math ML converter
LiDAR	light detection and ranging
LoRaWAN	long range wide area network
LPWAN	low power wide area network
M3IC	DARPA Magnetic, Miniaturized, and Monolithically Integrated Components program
MAA	Mathematical Association of America
MAC	medium access control
MAE	mean absolute error
MATCONT	Matlab package for numerical bifurcation analysis
MathML	Mathematical Markup Language (W3C standard)
MDS	multi-dimensional scaling
MEMS	micro-electrical mechanical systems
MGI	Materials Genome Initiative
MHD	magneto-hydrodynamics
MIP*	class of problems that can be verified through interactions with entangled quantum provers
ML	machine learning
MLP	mathematical language processing
MLP	multi-layer perceptron
MML	NIST Material Measurement Laboratory
MOLSSI	Molecular Science Software Institute
MOS	Magnus, Oberhettinger, and Soni
MPE	mean percent error
MPS	matrix product states
MPI	Message Passing Interface
MR	mixed reality
MRAM	magneto-resistive random-access memory
MRI	magnetic resonance imaging
MS	mass spectrometry
MSGI	NSF Mathematical Sciences Graduate Internship program
MSRI	Mathematical Sciences Research Institute (Berkeley)
muMAG	Micromagnetic Activity Group
nanoHUB	Web portal for nanotechnology research at https://nanohub.org/
NATO	North Atlantic Treaty Organization
NBS	National Bureau of Standards (former name of NIST)
NCNR	National Center for Neutron Research
NDA	Network Data Analytics (workshop)
NEI	National Emissions Inventory
NISQ	noisy intermediate-scale quantum
NIST	National Institute of Standards and Technology
NISTIR	NIST Internal Report
NITRD	Networking and Information Technology Research and Development
NLP	natural language processing
nm	nanometer
NMI	national metrology institute
NN	neural network
NRC	National Research Council
NSF	National Science Foundation
NYU	New York University

OCR	optical character recognition
ODE	ordinary differential equation
OECD	Organization for Economic Co-operation and Development
OOF	Object-Oriented Finite Elements (software)
OOF3D	3D version of OOF
OOMMF	Object-Oriented Micromagnetic Modeling Framework (software)
OP	orthogonal polynomials
OPSF	orthogonal polynomials and special functions
OS	photoreceptor outer segments
OSA	Optical Society of America
P2C	presentation to computation conversion
PAML	physics-assisted machine learning
PCR	polymerase chain reaction
PDE	partial differential equation
PH	Powell's hybrid method
PHY	physical layers (in a network)
PML	NIST Physical Measurement Laboratory
PPLN	periodically poled lithium nitrate
PQC	post-quantum cryptography
PREP	NIST Professional Research Education Program
PT	Painlevé transcendents
QA	quantum annealing
QAO	quantum adiabatic optimization
QAOA	quantum approximate optimization algorithm
QD	quantum dot
QDPD	Quaternion-based Dissipative Particle Dynamics simulation code
QIP	quantum information processing
QoS	quality of service
qPCR	quantitative polymerase chain reaction
QuICS	UMD-NIST Joint Center for Quantum Information and Computer Science
R&D	research and development
RAN	radio access network
RAVEN	IARPA Rapid Analysis of Various Emerging Nanoelectronics program
RBC	ray-based classification
RdCVF	rod-derived cone viability factor
RE	class of problems that are no harder than the halting problem
RGM	reduced gradient method
RGTM	research grade test material
RMSE	root mean square error
RNA	ribonucleic acid
RP	retinitis pigmentosa
RRM	radio resource management
RSA	Rivest-Shamir-Adelman public key cryptographic algorithm
RWC	region of wireless coexistence
SARS	severe acute respiratory syndrome
SARS-CoV-2	the virus that causes the respiratory disease coronavirus 19
SAVG	an internal HPCVG visualization file format
SDE	stochastic differential equation
SED	ITL Statistical Engineering Division
SEM	scanning electron microscope
sfp	symmetric free parameters
SHIP	Summer High School Internship Program
SIL	short iterative Lanczos
SIAM	Society for Industrial and Applied Mathematics
SIS	susceptible-infectious-susceptible
SIGGRAPH	ACM Special Interest Group on Graphics
SLA	service level agreement

SPDC	spontaneous parametric down conversion
SPIE	International Society for Optical Engineering
SRM	standard reference material
SSR-RoI	security risk reduction return on investment
STEM	science, technology, engineering, and mathematics
SUNY	State University of New York
SURF	NIST Student Undergraduate Research Fellowship program
SVD	singular value decomposition
SVM	support vector machine
SVOP	several variable orthogonal polynomials
SVP	NIST Student Volunteer Program
TDSE	time domain Schrodinger equation
TES	transition edge sensor
TPT	transition path theory
TQC	Theory of Quantum Computation (conference)
TSE	IEEE Transactions on Software Engineering
UCF	University of Central Florida
UMBC	University of Maryland Baltimore County
UMD	University of Maryland
UMIACS	University of Maryland Institute for Advanced Computer Studies
UNLV	University of Nevada Las Vegas
UQ	uncertainty quantification
URL	universal resource locator
USC	University of Southern California
UWB	ultrawide band
VaR	value at risk
VEMOS	Visual Explorer for Metric of Similarity (software)
VM	virtual machine
VNO	virtual network operator
VR	virtual reality
W3C	World Wide Web Consortium
WAS	Washington Academy of Sciences
WCE	wireless capsule endoscopy
WRF	weather research and forecasting
XR	extended reality
XSEDE	NSF eXtreme Science and Engineering Discovery Environment