NIST Special Publication 1275v3

NIST Conference Papers
Fiscal Year 2018
Volume 3:
Physical Measurement Laboratory
NIST Center for Neutron Research

Compiled and edited by:
Information Services Office

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1275v3

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

NIST Special Publication 1275v3

NIST Conference Papers
Fiscal Year 2018
Volume 3:
Physical Measurement Laboratory
NIST Center for Neutron Research

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

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1275v3

January 2022

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

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

National Institute of Standards and Technology Special Publication 1275v3
Natl. Inst. Stand. Technol. Spec. Publ. 1275v3, 377 **pages (January 2022)**
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1275v3

Foreword

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

This Special Publication represents the work of Physical Measurement Laboratory and NIST Center for Neutron Research researchers at professional conferences, as reported in Fiscal Year 2018.

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

Key words

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

Table of Contents

Measurement and accumulation of electric charge on a single dielectric particle trapped in air

Haesung Park[a] and Thomas W. LeBrun*[a]
[a]National Institute of Standards and Technology,
100 Bureau Drive, Gaithersburg, MD20899 USA

ABSTRACT

Normally occurring charges on small particles provide a means to control the motion of the particles. Using a piezoelectric transducer to launch microparticles into a trap, we can vary particle-surface interactions to transfer charge to the particle via contact electrification. This allows more detailed studies of contact electrification itself as well generation of higher charge states for precision measurements of force or nonlinear dynamics using electric field modulation. In practice, particles may be repeatedly landed on the substrate and relaunched during loading. This leads to charge transfer so that the net charge on the polystyrene (PS) particle becomes sufficient to allow electrostatic forcing to drive ballistic motion over a range of displacement two orders of magnitude greater than thermal fluctuations. An increase in charge from 1000 to 3000 electrons is demonstrated and the induced motion of the trapped particle is accurately described using simple classical mechanics in phase space.

Keywords: Optical levitation, contact electrification, electrostatic modulation, parametric force analysis, arbitrary force measurement, microparticles, optical tweezers, optical trapping, nonlinear dynamics

1. INTRODUCTION

The motion of an optically trapped particle is well approximated as a damped harmonic oscillator for small oscillations near the trap center. Once the trap stiffness (spring constant) is determined, the external force on the particle is easily measured using the induced displacement from the trap center $\left(F=-kx\right)$. Moreover, since the optical trap can be a highly ideal mechanical system comprised of light and the dielectric particle only, it helps us to expand our understanding on the nature of light and the system itself by measuring the optical force acting on the particle.[1] Therefore, optical traps have been used over a wide range of areas as a tool for highly sensitive measurement.[2–5]

In a previous study, we demonstrated parametric force analysis in which the ballistic motion of a trapped particle was recorded and used to calculate velocity and acceleration of the particle so that the force acting on the particle can be directly mapped out over a wide range with natural resonant frequency and damping measured near the trap center.[6] This doesn't require the conventional linear approximation, thus allowing direct measurement of arbitrary forces including the nonlinear regime. In parametric analysis, an electrostatic force can be employed to supply the external force creating ballistic motion because the dielectric particles are naturally charged during fabrication or handling. However the electrical charge on the individual polystyrene particles can vary significantly and be very difficult to predict or control.

In this paper, we demonstrated accumulation of electrical charges on the polystyrene (PS) dielectric particles so that the selected PS particle can be used as a probe to map out arbitrary (non-linear) forces over a wider range of the particle motion (i.e. unknown force field) without increasing the strength of applied electric field. This will allow in-depth understanding of the behavior of the generalized mechanical oscillator realized by the PS particle trapped in air beyond the linear regime.

*Corresponding author: lebrun@nist.gov

LeBrun, Thomas; Park, Haesung.
"Measurement and accumulation of electric charge on a single dielectric particle trapped in air."
Paper presented at SPIE OptoWest, San Francisco, CA, United States. February 13, 2016 - February 18, 2016.

2. THEORY

The motion of a trapped particle is well described by the Langevin equation. The details and application can be found in the literature.[6–8] Briefly, the motion of a particle is approximated as a harmonic (linear) oscillator near the trap center in one dimension,

$$\ddot{x}(t)+\beta\dot{x}(t)+\omega_o^2 x(t)=\Lambda\xi(t) \tag{1}$$

where $\omega_o\left(=\sqrt{k/m}\right)$ and $\beta\left(=\gamma/m\right)$ are the natural frequency and viscous damping for the particle (mass, *m*) with the stiffness k in a surrounding medium at temperature T with the coefficient of friction γ. Since the Brownian stochastic force, given by $\Lambda\xi(t)$ where $\Lambda=\sqrt{2k_BT\beta/m}$ with Boltzmann constant k_B, is random with zero mean force $\langle\xi(t)\rangle=0$, the time-averaged trajectory can in this case be treated as a simple harmonic oscillator without an external stochastic force i.e., $\ddot{x}(t)+\beta\dot{x}(t)+\omega_o^2 x(t)\cong 0$. For the simple harmonic oscillator, the natural frequency and damping are easily measured using the transient response which is the displacement response of the system to an applied external force.

In our previous study, we analyzed the transient (step) response of a trapped particle based on the ideal step response of a harmonic oscillator including additional terms for phase (Φ_{add}) and QPD offset (X_{off}). A nonlinear least square fit of the trajectory to the following equation is used,

$$x(t)=X_o\left(1-\frac{e^{-\varsigma\omega_o t}}{\sqrt{1-\varsigma^2}}\sin(\omega_o\sqrt{1-\varsigma^2}\,t+\phi+\Phi_{add})\right)+X_{off} \tag{2}$$

where $\phi\equiv\tan^{-1}\left(\sqrt{1-\varsigma^2}\,/\varsigma\right)$, and the damping ratio $\varsigma=\beta/\beta_{critical}$ is used to represent the motion relative to critically damped ($\beta_{critical}=2\omega_o$) response (i.e. the fastest response without overshoot).[9] Once ω_o, β, and X_o are measured by fitting the trajectory to the above form, the applied electrostatic force $\left(F_{es}=q\vec{E}=qV/d\right)$ balanced with optical restoring force $\left(F_{opt}=-kX_o\right)$ is easily calculated. Thus, the corresponding electrical charge on the trapped particle is measured using, $q=kX_o d/V$.[6] (The mass of the trapped particle is separately measured from the video microscope data using size and density.[10]) Applying the parametric force analysis, the optical forces are directly acquired from its equation of motion with a general force term without using linear approximation ($F_{res}=\omega_o^2 x$), i.e., $\ddot{x}(t)+\beta\dot{x}(t)+F_{res}\cong 0$ as rearranged in equation (3).

$$F_{res}\cong\ddot{x}(t)+\beta\dot{x}(t) \tag{3}$$

3. EXPERIMENTAL SETUP

The experimental setup used in this study is adopted from our previous experiment.[6,10] In many optical trap studies in gaseous media, a nebulizer is used to generate volatile liquid droplets which contain solid inclusions of

LeBrun, Thomas; Park, Haesung.
"Measurement and accumulation of electric charge on a single dielectric particle trapped in air."
Paper presented at SPIE OptoWest, San Francisco, CA, United States. February 13, 2016 - February 18, 2016.

nano/microparticles.[11,12] This approach is very convenient, especially for nanoparticle studies, however cannot provide the selectivity to introduce a specific particle into the optical trap. Also, the trapped particle cannot be tracked after exiting the trap or re-trapped. Here, a piezoelectric transducer is employed to load a selected particle over repeated cycles of loading and landing while the particle is triboelectrically charged as a result of contact electrification. The details of the piezoelectric launcher can be found in the literature.[6] Briefly, a ring-type piezoelectric transducer (PZT) is used to release the particle from the substrate, and the particles are imaged through the center hole of PZT. Once the particles are randomly applied on the glass substrate, the particles can be jiggled on the substrate under resonant PZT pulsing. Each excitation pulse consists of tens to hundreds of periods of a 64 kHz square wave. (Higher charges can lead to increased surface adhesion requiring greater excitation for release.) Inside the resulting fountain-like excitation region, the particles undergo random walks, and the PS particles become more highly charged through the contact electrification with the glass substrate or each other. This charging process can be repeated until the charge becomes sufficient for the intended purpose while the entire process is monitored with CCD camera.

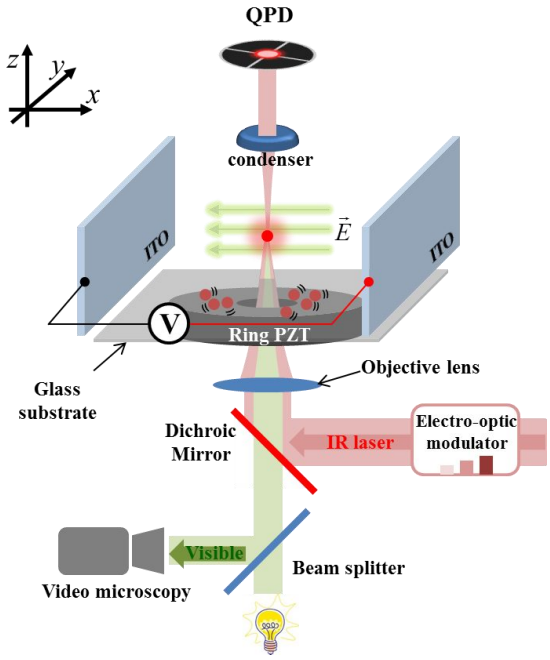


Figure 1 Schematics for the experimental setup: a ring type piezoelectric transducer (PZT) was used to release the polystyrene particles in a repeatable manner while the particles are monitored through video microscopy for the calibration of the quadrant cell photodetector (QPD) and tracking of the selected particle. Two parallel ITO-coated plates were used to generate a uniform electric field to drive ballistic motion of the charged particle.

Once the particles are released from the substrate, near-infrared (Nd:YVO_4, 1064 nm) light delivered through an objective lens (long working distance NIR corrected objective, Mitutoyo, 20X, NA = 0.4, WD = 20 mm) lifts the particle to the trapping volume. Otherwise the particles are returned to the surface. This process can be repeated if the desired particle is not delivered or the electrical charges are insufficient. The trapped particle can be landed by reducing the optical power (radiation pressure) so that gravity brings the particle to the glass substrate in a controlled manner. An electro-optic modulator is used to control the optical power. The trapped particle is protected by a glass cell consisting of five windows with two parallel conducting windows facing each other. (The bottom side is covered with a PZT holder.) The electric field is generated from two indium tin oxide (ITO) coated plates (15 mm × 15 mm) connected to the voltage amplifier with a function generator which is synchronously recorded with the particle trajectory through the data

acquisition (DAQ) system. The separation between the two plates is 10 mm. To avoid fringe fields, the particle was aligned to the center of the cell using the precision manual *xy* linear translational stage and motorized vertical *z*-axis stage, as shown in **figure 1**. The positional detection was performed with quadrant cell photodetector (QPD) and bright field video microscope for simultaneous QPD calibration.

4. RESULTS

When the microparticles are detached from the surface using the ultrasonic vibration, they acquire electrical charges from the glass substrate. This charge transfer through contact and separation is known as contact electrification (CE).[13] It is widely present in nature and has been used for consumer electronics such as photocopying and laser printing. Since polystyrene tends to acquire electrical charge from glass, PS particles can become highly charged (negative) through the repeated loading procedure.

To demonstrate charge accumulation on the selected PS particle, we measured electrical charges on the particle and then landed it by reducing the optical power while the particle is monitored through the CCD camera in real-time. The cycle of loading and landing is then repeated as desired until the particle escapes from the field of view or adheres to the substrate due to the increased electrostatic force. Once the particle is released and trapped stably, the electrical charges on the trapped particle are easily measured. The uncertainty of the charge measurement here is a few percent, but single electron resolution has been demonstrated.[14,15] **Figure 2** shows the electrical charge measured during 130 repeated cycles of the loading process leading to contact electrification. The initially measured electrical charge is seen to increase about three fold from the initial value of 1000 electrons (All charges here are reported in units of the elementary charge e, which is positive).

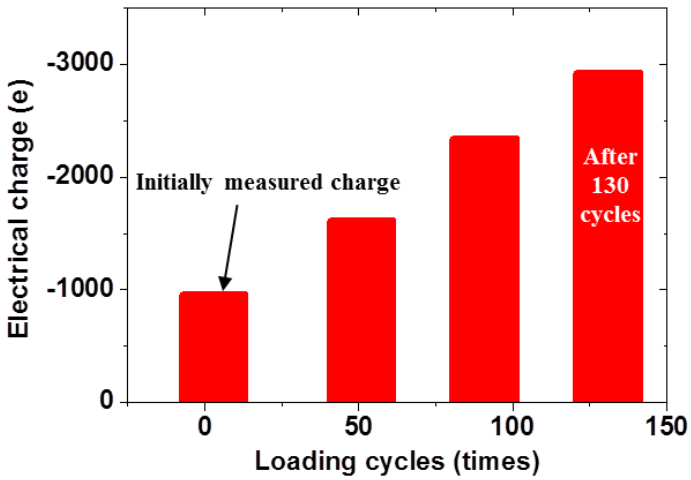


Figure 2 Electrical charge on the selected polystyrene (PS) particle was measured before and after the 130 cycles of loading procedure. The repeated loading makes the particle more negatively charged, agreeing with previous results showing PS particles to become more negatively charged after the contact with a glass substrate.[13] The charge was measured from the transient response of the trapped particle to the applied electrostatic force.

As we measured the natural frequency, damping, and induced displacement from the transient response, the force acting on the particle can now be estimated using the parametric force analysis. Based on the measured trajectory, the

LeBrun, Thomas; Park, Haesung.
"Measurement and accumulation of electric charge on a single dielectric particle trapped in air."
Paper presented at SPIE OptoWest, San Francisco, CA, United States. February 13, 2016 - February 18, 2016.

numerically calculated velocity and acceleration are used to calculate the resulting force from equation (3). The results are shown in **figure 3**. An electrostatic field of 40 V/mm is used to draw the particle approximately 300 nm from the center, corresponding to the initial position indicated by the red dot. It is then released, but the field does not turn off instantly. The vertical line rising from the initial represents the electrostatic field falling to zero. The estimated force quickly rises from zero (force balance) to a value corresponding to the optical force. The subsequent evolution of the line traces out the optical force curve. At higher initial displacements the estimated force in this one dimensional case diverges slightly from ideal behavior (perfect retrace) because off-axis excitation increases.

Figure 3 (a) and (b) show the estimated force acting on the same particle before and after using contact electrification to increase the charge by a factor of 3 (130 loading cycles). As a result the particle was displaced approximately three times further, mapping the optical restoring force acting on the particle over a wider range (from 300 nm to 900 nm) under the same applied electric field.

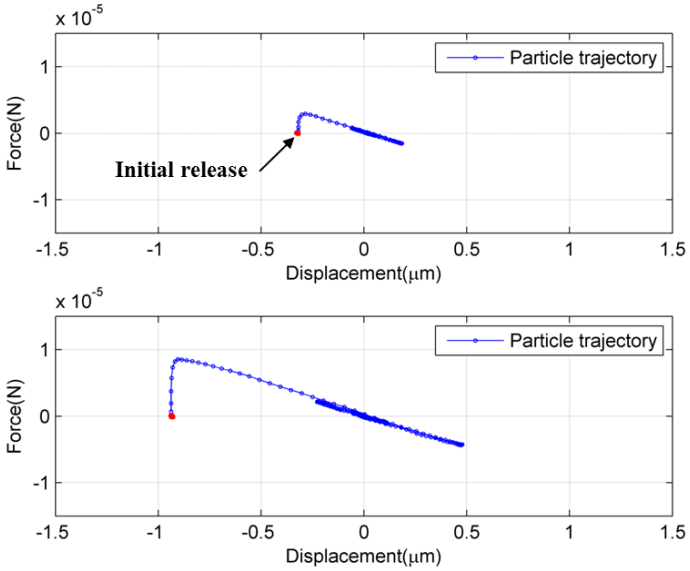


Figure 3 shows parametric force curves of the same particle (a) before and (b) after the contact electrification using the same strength of electrostatic field. As the particle becomes more charged, the same electric field generates more electrostatic force so that the induced displacement becomes three times wider than the initial range of motion. The red dots represent the initial position under the electrostatic force whereas the blue force curves are the force curve constructed from the retracted trajectories of the particle after the electric field is eliminated.

To describe the state of our dynamic system more completely in terms of classical mechanics, the measured phase space trajectory is shown. For the one dimensional dynamic system, the state of the system is determined by the values of position (x) and velocity ($\dot{x}$). Plotting the position and velocity of the trapped particle as function of time, the parametric curve on the phase plane describes the evolution of the trapped particle from the point of release (red dot). As shown in **figure 4**, the trapped PS particle shows the nearly ideal behavior of damped harmonic oscillator. After release, the restoring force draws the particle back to the origin while Stokes damping dissipates the kinetic energy. The trajectory converges to the trap center (blue line with closed circle), ultimately becoming dominated by Brownian motion (green line) as shown in the inset. The deviations in the force curve from linearity as shown in figure 3 are slight, so demonstrating the resulting nonlinear dynamics requires comparison to simulations which space does not permit here.

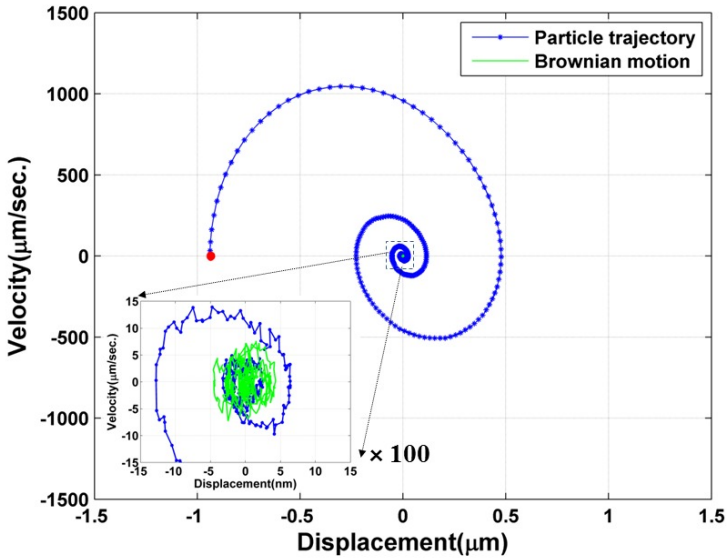


Figure 4 The measured phase space diagram for the trapped PS particle shows the ideal behavior of a damped oscillator. The "turning off" of the electrostatic force observed in figure 3 is not seen here (that is the points fall under the red dot) because the particle doesn't move appreciably during that time. At long times, the ballistic motion is reduced through Stokes damping of air, and Brownian motion dominates (green line).

5. SUMMARY

In this paper, we have demonstrated that the electrical charges on a specific polystyrene particle can be increased through repeated loading processes. As the piezoelectric transducer drives a particle adhered to a surface, the inertial force detaches the particle so that electrical charges are transferred to the particle when it is released from the glass substrate. Through numerous cycles of re-loading, the particle accumulates progressively more electrical charge and becomes a more useful probe of the trap, so that the optical restoring force acting on the particle can be deduced from its dynamics using a parametric force analysis. This allows us to map out arbitrary forces acting on the particle within the trapping volume and illuminate the important yet poorly-understood electrification process itself.

REFERENCES

[1] Kyrsting, A., Bendix, P. M.., Oddershede, L. B., "Mapping 3D focal intensity exposes the stable trapping positions of single nanoparticles.," Nano Lett. **13**(1), 31–35 (2013).
[2] Svoboda, K., Schmidt, C. F., Schnapp, B. J.., Block, S. M., "Direct observation of kinesin stepping by optical trapping interferometry.," Nature **365**(6448), 721–727 (1993).
[3] Li, T.., Raizen, M. G., "Brownian motion at short time scales," Ann. Phys. **525**(4), 281–295 (2013).
[4] Mehta, A. D., "Single-Molecule Biomechanics with Optical Methods," Science (80-.). **283**(5408), 1689–1695 (1999).
[5] Arvanitaki, A.., Geraci, A. A., "Detecting High-Frequency Gravitational Waves with Optically Levitated Sensors," Phys. Rev. Lett. **110**(7), 071105 (2013).
[6] Park, H.., LeBrun, T. W., "Parametric Force Analysis for Measurement of Arbitrary Optical Forces on Particles Trapped in Air or Vacuum," ACS Photonics **2**(10), 1451–1459, American Chemical Society (2015).
[7] Thornton, S. T.., Marion, J. B., Classical Dynamics of Particles and Systems, 5th ed., Brooks/Cole (2003).

[8] Le Gall, A., Perronet, K., Dulin, D., Villing, A., Bouyer, P., Visscher, K.., Westbrook, N., "Simultaneous calibration of optical tweezers spring constant and position detector response.," Opt. Express **18**(25), 26469–26474, OSA (2010).
[9] Nise, N. S., Control Systems Engineering, Wiley (2003).
[10] Park, H.., LeBrun, T. W., "Measuring forces and dynamics for optically levitated 20μm PS particles in air using electrostatic modulation," SPIE Nanosci. + Eng., K. Dholakia and G. C. Spalding, Eds., 95480E, International Society for Optics and Photonics (2015).
[11] Summers, M. D., Burnham, D. R.., McGloin, D., "Trapping solid aerosols with optical tweezers: A comparison between gas and liquid phase optical traps," Opt. Express **16**(11), 7739–7747, OSA (2008).
[12] Anand, S., Nylk, J., Neale, S. L., Dodds, C., Grant, S., Ismail, M. H., Reboud, J., Cooper, J. M.., McGloin, D., "Aerosol droplet optical trap loading using surface acoustic wave nebulization.," Opt. Express **21**(25), 30148–30155, Optical Society of America (2013).
[13] McCarty, L. S.., Whitesides, G. M., "Electrostatic charging due to separation of ions at interfaces: contact electrification of ionic electrets.," Angew. Chem. Int. Ed. Engl. **47**(12), 2188–2207 (2008).
[14] Moore, D. C., Rider, A. D.., Gratta, G., "Search for Millicharged Particles Using Optically Levitated Microspheres," Phys. Rev. Lett. **113**(25), 251801 (2014).
[15] Beunis, F., Strubbe, F., Neyts, K.., Petrov, D., "Beyond Millikan: The Dynamics of Charging Events on Individual Colloidal Particles," Phys. Rev. Lett. **108**(1), 016101 (2012).

LeBrun, Thomas; Park, Haesung.
"Measurement and accumulation of electric charge on a single dielectric particle trapped in air."
Paper presented at SPIE OptoWest, San Francisco, CA, United States. February 13, 2016 - February 18, 2016.

NANOSCALE TUNING FORK CAVITY OPTOMECHANICAL TRANSDUCERS WITH DESIGN-ENABLED FREQUENCY TUNING AND TEMPERATURE COMPENSATION

Rui Zhang[1], Robert Ilic[2], Yuxiang Liu[1], and Vladimir Aksyuk[2]

[1]Department of Mechanical Engineering, Worcester Polytechnic Institute, USA
[2]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, USA

ABSTRACT

In this work, we design, fabricate and characterize monolithic, nanoscale Si_3N_4 tuning fork cavity optomechanical transducers with design enabled tuning of mechanical resonant frequencies and passive temperature compensation. Both frequency tuning and temperature compensation are achieved by the design of a nonlinear mechanical clamp. The compensation reduces the temperature sensitivity of frequency by ≈ 60 times, achieving fractional frequency sensitivity of $(3.2 \pm 0.4)\cdot 10^{-6}$ K^{-1}. The design simultaneously increases the stress in the tuning fork by more than a factor of two relative to the residual stress, and makes the resulting high resonator frequency insensitive to the residual stress and temperature variations. These stable resonators may find uses for stable transduction of force and motion in MEMS devices via frequency readout.

INTRODUCTION

Many mechanical sensors, such as high performance inertial sensors, demand integrated mechanical motion detection with low noise, high bias stability, large dynamic ranges, and high bandwidths. Monitoring the resonant frequency change of a high-quality mechanical resonator is a widely accepted approach for motion detection. This approach is immune to low-frequency electronic noise and bias drift, and both a high mechanical frequency (f_m) and a high mechanical quality factor (Q_m) are preferred since high f_m can provide high transduction bandwidth/temporal resolution and the force sensitivity scales as $1/(f_m^{0.5}Q_m^{0.5})$ [1]. Nanomechanical resonators made of high intrinsic tensile stress silicon nitride (Si_3N_4) have shown great potential for such sensing applications [2] because of their high mechanical quality factors, high frequencies, and facile fabrication. Recently developed nanoscale Si_3N_4 tuning fork resonators [3] achieve very large frequency-quality factor (f_mQ_m) products ($Q_m \approx 2.2\times10^5$ at $f_m \approx$ 29 MHz) through a design-enabled stress tuning approach, and open unprecedented opportunities for on-chip motion metrology and sensing. However, the frequency of such tuning forks is determined by both designed geometry and the intrinsic stress (σ_0), and σ_0 varies with different fabrication processes [4]. In addition, the performance of the tuning fork mechanical resonator can be influenced by ambient environmental factors, especially the temperature [5]. For example, temperature fluctuations can influence the tensile stress in a doubly-clamped Si_3N_4 tuning fork, due to the difference of coefficient of thermal expansion (CTE) between the Si_3N_4 device layer and the Si substrate. Thermal stress fluctuations in turn influence the thermal stability of the resonant frequency [6]. The majority of approaches made to compensate temperature-induced frequency variation rely on matching different materials [7] or active temperature control [8].

Here we design, fabricate, and characterize a monolithic nonlinear mechanical clamp for single-layer Si_3N_4 tuning fork nanomechanical resonators that renders the natural frequencies insensitive to temperature fluctuations by a passive, design-enabled compensation method. In addition, by introducing a reference length

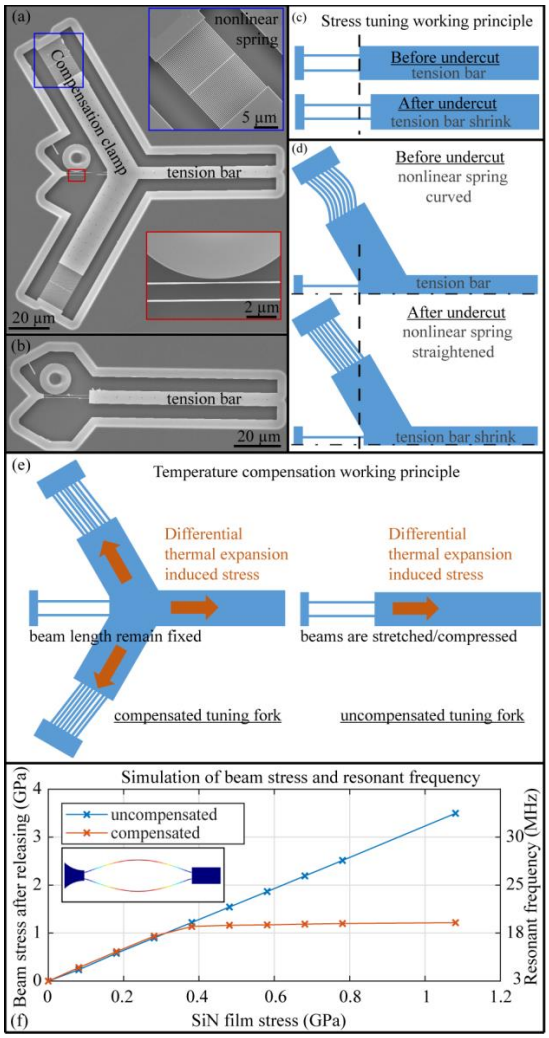


Figure 1: (a) SEM of a tuning fork with temperature compensation and integrated photonic readout. (Top inset) Zoom-in of straightened nonlinear springs. (Bottom inset) Coupling region between tuning fork beam and microdisk optical resonator. (b) SEM of tuning fork without temperature compensation. (c) Working principle of stress tuning in uncompensated tuning fork. (d) Working principle of stress tuning in compensated tuning fork. (e) Working principle of temperature compensation. (f) Simulation of beam stress and device frequency with different Si_3N_4 film stress. The inset shows the simulated tuning fork in-plane squeezing mode shape with exaggerated deformation.

Aksyuk, Vladimir; An, Sang; Centrone, Andrea; Chae, Jungseok; Holland, Glenn; Zou, Jie.
"Nanoscale tuning fork cavity optomechanical transducers with design-enabled frequency tuning and temperature compensation."
Paper presented at Hilton Head 2016 Workshop, Hilton Head Island, SC, United States. June 5, 2016 - June 9, 2016.

scale via the device layout, this clamp design can achieve the same designed resonator strain and resonant frequency over a wide range of intrinsic stress values and enables geometry-determined on-chip stress tuning capability. Different temperature compensation strengths including under-compensation, proper compensation, and over-compensation can be achieved by varying the design and are demonstrated in the experiments. The device with the best temperature compensation has a temperature sensitivity of -54 Hz/K ± 6 Hz/K with $f_m \approx 16.5$ MHz. The tuning fork without temperature compensation has a temperature sensitivity of 5360 Hz/K ± 170 Hz/K with $f_m \approx 27.76$ MHz. All reported statistical uncertainties are 95 % confidence ranges unless noted otherwise. The nonlinear mechanical clamp can be adapted for a wide range of doubly-clamped nanomechanical resonators to reduce temperature sensitivity.

Scanning electron microscopy (SEM) images of temperature compensated and uncompensated tuning forks are shown in Figures 1(a) and (b), respectively. The uncompensated tuning fork consists of two parallel cantilever beams, and is doubly clamped with a long tension bar on the right. By comparison, the compensated tuning fork has two additional inclined clamps, resulting in a Y-shaped clamp at the right end of the tuning fork. Each of the inclined clamps is attached to the supporting structure by an array of nonlinear springs. All the devices are fabricated in a 250 nm thick stoichiometric Si_3N_4 layer, which is grown by low-pressure chemical vapor deposition (LPCVD) on a 3 μm thick silicon dioxide (SiO_2) layer on a silicon substrate. We first used electron-beam (E-beam) lithography to define the geometry in a layer of positive tone E-beam resist. The pattern was then transferred to the Si_3N_4 film by a $CHF_3/Ar/O_2$ inductively-coupled plasma reactive ion etch (ICP RIE). Device fabrication is finished by a buffered oxide etch (BOE) undercut, which selectively removes the SiO_2 to release the device with minimal etching of Si_3N_4.

The nominal length and width of the tuning fork beams are 20 μm and 150 nm, respectively, in all the fabricated devices. The tension bar on the right-hand side of the tuning fork is 80 μm long. In the uncompensated device, the tension bar shrinks after device release due to the redistribution of the initially uniform Si_3N_4 stress, such that the stress in the tuning fork beam increases, as shown in Figure 1(c). Because the length of the tension bar determines the final stress in the tuning fork after undercut, one can tune the fork mechanical frequency simply by the design of the tension bar. This design-enabled tuning has been demonstrated in our previous work to achieve an up-to-three-fold stress increase, compared with the intrinsic Si_3N_4 film stress, and 1.5 times increase in the fork fundamental frequency compared with the fork without the tension bar [3].

In the temperature-compensated device (shown in Figure 1(a)), the Y-shaped clamp includes two symmetric inclined clamps and a tension bar. Each of the two symmetric, inclined compensation clamps has an array of nonlinear springs close to one end. The springs are curved before undercut with the shape defined by one period of a cosine function (0.48 μm) × cos(x/(20 μm) × 2π), where 0.48 μm is the modulation of the cosine geometry and x is the position in micrometers along the longitudinal direction. The symmetric inclined clamps and the curved springs allow the tuning fork beam stress to be insensitive to the intrinsic Si_3N_4 stress, while retaining the stress tuning capability, as described below. During the release process (Figure 1(d)), the tension bar first shrinks, as in the uncompensated device, stretching the tuning forks and straightening the curved springs in the inclined clamps. As a result, the tuning fork beam stress and frequency are tuned by the tension bar design, until the curved springs are fully straightened. If the intrinsic Si_3N_4 film stress is large enough to fully straighten the nonlinear springs, the stiffness of the straight nonlinear springs increases significantly compared with that of curved springs. This is because the spring deformations change from curved beam bending to straight beam stretching. The inclined clamps then provide an effective balancing force in the opposite direction of the stretching force produced by the tension bar. As a result, the right-hand end-point of the tuning fork beams is fixed and hence the tuning fork beam stress as well as mechanical frequency become insensitive to the intrinsic Si_3N_4 film stress.

The stress insensitivity of the frequency enabled by the clamp design is verified by a proof-of-concept Finite-Element-Method (FEM) simulation. The simulated final beam stress and mechanical frequency at different Si_3N_4 intrinsic stress are shown in Figure 1(f). In the uncompensated device, the beam stress is linearly proportional to the intrinsic stress. In the compensated device, the beam stress in the tuning fork increases similarly for small intrinsic stress, before the nonlinear springs are fully straightened. However, the beam stress is almost constant for the intrinsic stress above a threshold value at which the nonlinear springs become straight and stiff. The final beam stress in the compensated device is determined by the clamp geometry (the threshold displacement required to straighten the nonlinear spring) and is insensitive to the exact value of the Si_3N_4 intrinsic film stress.

The inclined clamps also enable the tuning fork frequency to be insensitive to temperature changes. The working principle of temperature compensation is shown in Figure 1(e). When the temperature changes, the differential thermal expansion due to CTE mismatch between the Si_3N_4 device layer and the silicon substrate results in thermal stress in the uncompensated devices. Specifically, the Si substrate expands more than the Si_3N_4 device layer when the temperature increases, increasing the beam stress and in turn the resonant frequency of the tuning fork. However, in the compensated devices, the temperature-induced stress from the tension bar is balanced by that from the inclined clamps. Thus, the tuning fork beam anchor remains fixed regardless of temperature variation, providing a temperature independent mechanical frequency. The compensation strength can also be tuned by the clamp geometry. For example, the temperature induced force from the compensation clamps can be tuned by the width of the nonlinear springs. The temperature induced force from the tension bar can be tuned by the width of the tension bar. As a result, small, proper, and large ratios of nonlinear-spring to tension-bar width can produce under-compensated, properly compensated, and over-compensated tuning forks, respectively. FEM simulation shows that a device with a 3.2 μm wide tension bar and 200 nm wide nonlinear springs can have almost zero temperature sensitivity. Thus, to investigate the influence of the width ratio between the nonlinear spring and the tension bar on the compensation strength, we fabricated devices with the nominal tension bar width varying from 2.8 μm to 3.7 μm and nonlinear spring widths varied from 194 nm to 211 nm.

The mechanical frequencies of the tuning forks are measured through a near-field cavity-optomechanical readout [9]. Each tuning fork is near-field coupled across an ≈ 150 nm gap to a Si_3N_4 microdisk optical cavity. The disk has a nominal diameter of 14 μm and supports whispering-gallery (WGM) optical modes. When the light wavelength is tuned to the shoulder of a microdisk optical resonance, the tuning fork beam motion can modulate the optical resonance and hence change the transmitted light intensity. Then the output light signal can be transduced into electric signal by a photodetector for post processing. The optomechanical coupling parameter $g_{om}/2\pi$ between the tuning fork in-plane squeezing mechanical mode and the microdisk WGM optical mode is calculated to be ≈ 140 MHz/nm, which indicates the optical resonance frequency shift per unit modal displacement of the

Aksyuk, Vladimir; An, Sang; Centrone, Andrea; Chae, Jungseok; Holland, Glenn; Zou, Jie.
"Nanoscale tuning fork cavity optomechanical transducers with design-enabled frequency tuning and temperature compensation."
Paper presented at Hilton Head 2016 Workshop, Hilton Head Island, SC, United States. June 5, 2016 - June 9, 2016.

mechanical resonator. This near-field optical readout yields ≈ 1 fm/Hz$^{0.5}$ displacement resolution [9] and enables a monolithic, robust, and fiber-pigtailed sensor technology [10].

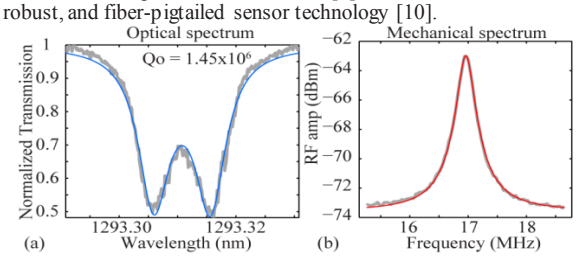


Figure 2. Experimentally measured (a) optical spectrum (b) mechanical spectrum. Colored lines are theory fits.

To measure the mechanical frequencies, light was emitted from a 1300 nm wavelength band tunable laser and evanescently coupled to the optical microdisk resonator through a fiber taper helix probe [11]. The transmitted light intensity was measured by a photo detector (PD), with the output split into two channels. One channel connected to a data acquisition (DAQ) board for swept-wavelength spectroscopy of the optical cavity modes (Figure 2(a)). The other channel connected to an Electrical Spectrum Analyzer (ESA) to measure the mechanical frequency (Figure 2(b)). Since the optomechanical transduction noise is well below the random motion of the tuning fork driven by thermomechanical noise [12], the mechanical frequencies were measured from the thermal motion spectra of the tuning forks without external excitation. The temperature was stably controlled within ± 0.1 K by a ceramic heater with a proportional–integral–derivative (PID) controller. The device was enclosed inside an acrylic chamber, with desiccant boxes placed in the chamber and constant desiccated air flowing through the chamber, to minimize the influence from moisture on the devices. At room temperature the frequencies of the first-order squeezing mode were ≈ 27.7 MHz and ≈ 16.5 MHz for the uncompensated and compensated tuning forks, respectively.

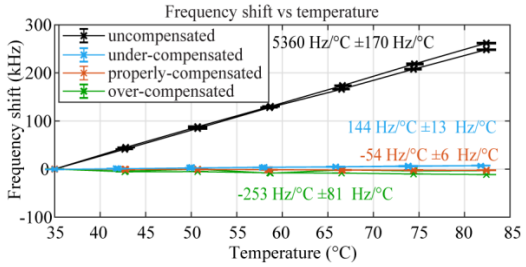


Figure 3. Experimentally measured temperature induced frequency shift. The frequency uncertainties are determined by nonlinear least squares fits to the data. The uncertainties for the temperature sensitivity are determined by the linear least squares fits.

To verify the temperature compensation effect, the mechanical frequency shifts of different devices over the temperature range from 35 °C to 82 °C were measured and shown in Figure 3. The resonant frequency of uncompensated device increased linearly with temperature at the rate of ≈ 5360 Hz/K ±170 Hz/K. For a beam vibration dominated by the tensile stress, the fundamental mode resonant frequency f_m as a function of temperature T can be described as [6]

$$f_m(T) = \frac{1}{2L}\sqrt{\frac{\sigma - AE(\alpha_{beam}-\alpha_{sub})(T-T_0)}{\rho}} \quad (1)$$

where L is the beam length, σ is the tensile stress along the beam at reference temperature, ρ is the density of the beam, E is the Young's modulus of the beam material, α_{beam} and α_{sub} are the CTE of the beam and substrate, respectively, and T_0 is the reference temperature. For the tuning fork with a tension bar, the stress is amplified by an amplification ratio $A = \sigma/\sigma_0$, and the stress variation due to temperature is similarly amplified. As a result, for the tuning fork with stress tuning, equation (1) is modified,

$$f_m(T) = \frac{1}{2L}\sqrt{\frac{\sigma - AE(\alpha_{beam}-\alpha_{sub})(T-T_0)}{\rho}} \quad (2)$$

in which σ is the amplified beam stress. Since the CTE of the Si substrate (≈ 2.6·10^{-6} K^{-1}) is larger than that of Si_3N_4 device layer (≈ 1.6·10^{-6} K^{-1}), the f_m increases as temperature increases. In addition, because $\sigma \approx 1.81$ GPa (calculated by FEM) is much larger than $E(\alpha_{beam}-\alpha_{sub})T_0$, the higher order terms of $f_m(T)$ can be neglected. Thus, f_m is approximately linearly-proportional to T which agrees with the experiment result. The temperature sensitivity can be then calculated as

$$S = \left.\frac{\partial f_m}{\partial T}\right|_{T=T_0} = -\frac{AE(\alpha_{beam}-\alpha_{sub})}{4L\sqrt{\rho\sigma}} \quad (3)$$

By using FEM-calculated values of $A \approx 3.12$, $E \approx 310$ GPa and $\rho \approx 3000$ kg/m^3 for the Si_3N_4 film, the temperature sensitivity is calculated to be ≈ 5200 Hz/K, which also matches the experiment.

For compensated devices, the resonant frequencies are also linearly-proportional to the temperature change but with much lower temperature sensitivity. This indicates that the nonlinear springs are fully deployed and the differential thermal strain is compensated by the nonlinear clamp design. The two inclined compensation clamps pull the center of the Y shaped structure in the direction opposite to that of the tension bar, balancing out the thermal stress induced motion. In addition, the strength of the compensation achieved by the Y-shaped clamp can be engineered by changing the width ratio between nonlinear springs and tension bar. With a proper width ratio, the temperature-induced movement of the center of Y-shaped clamp is such that the tuning fork does not experience a change in stress, which results in a perfect temperature compensation in the tuning fork. The thermal stress in the tension bar is dominant if the width ratio is too small, resulting in under-compensation, while a too large width ratio results in over-compensation.

As shown in Figure 3, three temperature compensation strengths are experimentally demonstrated using different nonlinear spring to tension-bar width ratios. Specifically, a device with nominally 177 nm wide nonlinear springs and a 3.14 µm wide tension bar (width ratio 0.056) has a temperature sensitivity of 144 Hz/K ± 13 Hz/K, indicating this device is under-compensated. A device with -253 Hz/K ± 81 Hz/K temperature sensitivity (over-compensated) has 230 nm wide nonlinear springs and a 3 µm wide tension bar (width ratio 0.077). The minimum temperature sensitivity -54 Hz/K ± 6 Hz/K (fractional frequency sensitivity of (3.2 ± 0.4)·10^{-6} K^{-1}) is measured from device with a 183 nm wide spring and a 2.96 µm wide tension bar (width ratio 0.062). Comparing to the uncompensated tuning fork which has 5360 Hz/K ± 170 Hz/K (fractional sensitivity (193 ± 6)·10^{-6} K^{-1}), the temperature sensitivity is reduced by a factor of ≈ 60 using the temperature compensation design. Temperature sensitivity uncertainties are determined from the linear fits.

Although our devices have not yet been tested directly as mechanical motion sensors, the displacement sensitivity of the resonance frequency is numerically and analytically estimated to be ≈ 51 kHz/nm for the uncompensated fork and ≈ 74 kHz/nm for the compensated fork based on the device dimension and resonant frequency measured in the experiments. These results confirm that the additional inclined clamps help to reduce the temperature induced frequency variability and the tuning fork transducer can be potentially used for force and displacement sensing with better stability.

Aksyuk, Vladimir; An, Sang; Centrone, Andrea; Chae, Jungseok; Holland, Glenn; Zou, Jie.
"Nanoscale tuning fork cavity optomechanical transducers with design-enabled frequency tuning and temperature compensation."
Paper presented at Hilton Head 2016 Workshop, Hilton Head Island, SC, United States. June 5, 2016 - June 9, 2016.

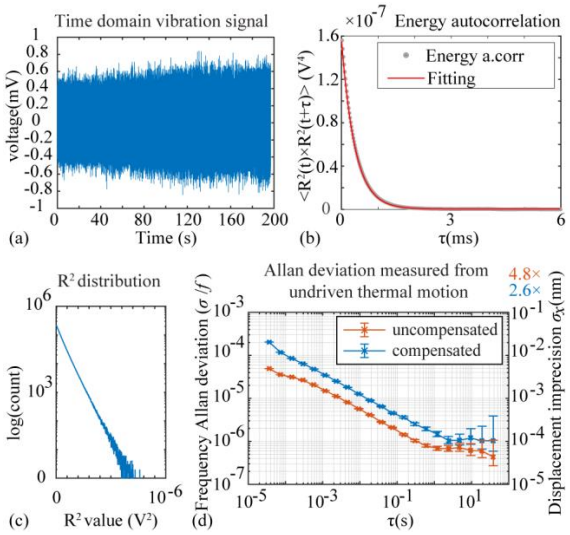


Figure 4. (a). Tuning fork time domain vibration signal. (b). Energy autocorrelation calculated from time domain signal. (c). R^2 distribution in log scale. (d). Experimentally measured frequency Allan deviation for compensated and uncompensated tuning forks, without excitation. Right-hand-side axis shows the corresponding calculated Allan deviation for a displacement sensing application with scale multipliers 4.8× (uncompensated) and 2.6× (compensated) due to different displacement sensitivity. The uncertainties for Allan deviation are determined by Chi distribution.

To further investigate the frequency stability of the tuning fork resonators, the time-domain vibration signal (quadrature, IQ, components at resonance) of the tuning fork thermal motion was also measured in vacuum using a lock-in amplifier. The I component of the uncompensated tuning fork time domain vibration signal is shown in Figure 4(a). Figure 4(b) shows the auto-correlation of its potential energy $Sa(\tau)=\langle R^2(t) \times R^2(t+\tau)\rangle$ where R is the signal amplitude. The calculated energy relaxation time $T_1 \approx 10^{-4}$ s indicates a sensor bandwidth exceeding 1 kHz. Figure 4(c) shows the distribution of R^2, which is a straight line in log scale indicating that, as expected, the energy obeys the Maxwell–Boltzmann distribution ($\propto \exp(-\alpha R^2/(2k_BT))$), where k_B is the Boltzmann constant and α is a constant. This indicates the linearity of the readout and of the system itself being mechanically linear and in thermodynamic equilibrium. From the time domain data, the frequency was measured as a function of time and the Allan deviation calculated, shown in Figure 4(d). The Allan deviation slope measured for the uncompensated tuning fork with 28 MHz mechanical frequency is $\approx 10^{-6}\,\mathrm{Hz}^{-0.5}$ below 1 s averaging, which is thermodynamically limited [12]. The relative bias stability is $\approx 10^{-6}$ above 1 s averaging. The Allan deviation measured from the compensated device is similar to that for the uncompensated one, indicating that the temperature compensation is achieved without increasing noise from other noise sources [12].

In conclusion, we have developed Si_3N_4 tuning fork cavity optomechanical transducers with nonlinear temperature compensation clamps. The clamp design allows the tuning fork mechanical frequency to be determined only by the geometry, regardless of the Si_3N_4 intrinsic film stress. Control and reduction of temperature sensitivity was experimentally demonstrated by tuning the design geometry. Temperature sensitivity ranges from -253 Hz/K ± 81 Hz/K (over-compensation) to 144 Hz/K ± 13 Hz/K (under-compensation) are achieved. The minimum temperature sensitivity is -54 Hz/K ± 6 Hz/K while that of an uncompensated tuning fork is 5360 Hz/K ± 170 Hz/K. The slope of measured thermodynamically-limited Allan deviation is $\approx 10^{-6}\,\mathrm{Hz}^{-0.5}$ below 1 s averaging and the relative bias stability is $\approx 10^{-6}$ above 1 s averaging. The temperature-compensated resonant transducer presented here demonstrates a promising new high-precision, low-drift displacement measurement approach that can enable a variety of precision MEMS sensor applications.

REFERENCES

[1] K. Y. Yasumura, T. D. Stowe, E. M. Chow, T. Pfafman, T. W. Kenny, B. C. Stipe, et al., "Quality factors in micron-and submicron-thick cantilevers," Microelectromechanical Systems, Journal of, vol. 9, pp. 117-125, 2000.

[2] A. Suhel, B. D. Hauer, T. S. Biswas, K. S. D. Beach, and J. P. Davis, "Dissipation mechanisms in thermomechanically driven silicon nitride nanostrings," Applied Physics Letters, vol. 100, p. 173111, 2012.

[3] R. Zhang, C. Ti, M. I. Davanço, Y. Ren, V. Aksyuk, Y. Liu, et al., "Integrated tuning fork nanocavity optomechanical transducers with high f_MQ_M product and stress-engineered frequency tuning," Applied Physics Letters, vol. 107, p. 131110, 2015.

[4] P. Temple-Boyer, C. Rossi, E. Saint-Etienne, and E. Scheid, "Residual stress in low pressure chemical vapor deposition SiNx films deposited from silane and ammonia," Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films, vol. 16, pp. 2003-2007, 1998.

[5] T. Kose, K. Azgin, and T. Akin, "Design and fabrication of a high performance resonant MEMS temperature sensor," Journal of Micromechanics and Microengineering, vol. 26, p. 045012, 2016.

[6] T. Larsen, S. Schmid, L. Grönberg, A. Niskanen, J. Hassel, S. Dohn, et al., "Ultrasensitive string-based temperature sensors," Applied Physics Letters, vol. 98, p. 121901, 2011.

[7] D. R. Myers, R. G. Azevedo, L. Chen, M. Mehregany, and A. P. Pisano, "Passive Substrate Temperature Compensation of Doubly Anchored Double-Ended Tuning Forks," Journal of Microelectromechanical Systems, vol. 21, pp. 1321-1328, 2012.

[8] J. C. Salvia, R. Melamud, S. A. Chandorkar, S. F. Lord, and T. W. Kenny, "Real-Time Temperature Compensation of MEMS Oscillators Using an Integrated Micro-Oven and a Phase-Locked Loop," Journal of Microelectromechanical Systems, vol. 19, pp. 192-201, 2010.

[9] Y. Liu, H. Miao, V. Aksyuk, and K. Srinivasan, "Wide cantilever stiffness range cavity optomechanical sensors for atomic force microscopy," Optics express, vol. 20, pp. 18268-18280, 2012.

[10] J. Chae, S. An, G. Ramer, V. Stavila, G. Holland, Y. Yoon, et al., "Nanophotonic Atomic Force Microscope Transducers Enable Chemical Composition and Thermal Conductivity Measurements at the Nanoscale," Nano Letters, vol. 17, pp. 5587-5594, 2017/09/13 2017.

[11] Y. Ren, R. Zhang, C. Ti, and Y. Liu, "Tapered optical fiber loops and helices for integrated photonic device characterization and microfluidic roller coasters," Optica, vol. 3, pp. 1205-1208, 2016.

[12] M. Sansa, E. Sage, E. C. Bullard, M. Gély, T. Alava, E. Colinet, et al., "Frequency fluctuations in silicon nanoresonators," Nature Nanotechnology, vol. 11, p. 552, 02/29/online 2016.

CONTACT

*V. Aksyuk, tel: +1-301-975-2867; vladimir.aksyuk@nist.gov

Aksyuk, Vladimir; An, Sang; Centrone, Andrea; Chae, Jungseok; Holland, Glenn; Zou, Jie.
"Nanoscale tuning fork cavity optomechanical transducers with design-enabled frequency tuning and temperature compensation."
Paper presented at Hilton Head 2016 Workshop, Hilton Head Island, SC, United States. June 5, 2016 - June 9, 2016.

Frequency shift mitigation in a cold-atom CPT clock

Xiaochi Liu[1], Eugene Ivanov[2], John Kitching[1], and Elizabeth A. Donley[1]
[1]Time and Frequency Division, NIST, Boulder, Colorado, USA
[2]School of Physics, the University of Western Australia, Perth, Australia

***Abstract* — An upgrade in the laser system for our cold-atom clock based on coherent population trapping is described that makes use of an electro-optic modulator to significantly improve the phase coherence of the interrogation spectrum relative to what was achieved with our previous phase-locked loop approach. With this improvement in phase coherence, we have observed a reduction in the light shift by over two orders of magnitude. We present the impact of this improvement on the performance of the clock, for which the light shift was previously dominated by the incoherent light in the laser spectrum. We also demonstrate a new optical setup to reduce the Doppler shift.**

Keywords—Phase-locked loop, coherent population trapping, light shift, AC Stark Shift, electro-optic modulator

I. Introduction

Systems producing phase-coherent optical signals separated by a few gigahertz are required for the development of various quantum-mechanical devices including those based on Coherent Population Trapping (CPT) [1]. A common technique for generating this light is to use Optical Phase-Locked Loops (OPLLs), which is the approach that we initially applied for our cold-atom CPT clock [2-4]. This technique enables a high degree of phase control, provided that both master and slave lasers have narrow linewidths and the control system has relatively high loop bandwidth [5]. However, it is also desirable to produce the bichromatic light with compact single-section DBR and DFB lasers, which are characterized by broad linewidths that are of the same order of magnitude as their frequency modulation bandwidth.

In the present work, we describe an upgrade to our interrogation spectrum achieved by generating the interrogation light with a fiber-coupled electro-optic phase modulator (EOM). We also present an initial evaluation of the impact of the improved spectrum on the light shift in our cold-atom CPT clock. For this clock, the light shift was previously dominated by phase noise in the CPT spectrum due to incomplete phase locking [4]. We show that the sensitivity to light shifts is much reduced with the improved spectrum.

II. OPLL versus EOM Light Source

In [2], we described a bichromatic light source based on two single-section distributed Bragg reflector lasers phase-locked across the 6.835 GHz hyperfine ground-state splitting of ^{87}Rb. With this approach, the fraction of total power in the coherent carrier of the beat note was limited to about 70 %. The lack of complete phase coherence is due to the slowness of the laser diode's frequency response. Single-section diode lasers typically exhibit a phase reversal in their frequency modulation response in the 0.5 MHz to 10 MHz range due to the crossover between thermally and carrier-density induced tuning regimes [6], which limits the maximum achievable bandwidth of the OPLL in our system to about 1 MHz. Since this is comparable to the linewidth of the free-running lasers, the degree of coherence between master and slave is limited with this approach.

To improve the coherence of the bichromatic light for the CPT excitation, we switched to a fiber-coupled phase EOM-based light source. Fiber-coupled phase EOMs are characterized by broad bandwidths (> 10 GHz) and low π phase-shift voltages. They are well suited for the generation of multi-frequency phase-coherent laser beams when they are driven by low-noise microwave signals. Fig. 1 shows the scheme of the CPT clock with the EOM-based light source.

The phase coherence between the two sidebands generated by the EOM is much higher than the OPLL based two-laser system. The beat-note spectrum of the EOM-based light source shows an improved coherence of ~99% -- a significant improvement over the 70 % achieved with the OPLL. A disadvantage of the EOM-based approach is that some of the optical power is in sidebands that do not contribute to the CPT signal but can nevertheless contribute to light shifts and noise.

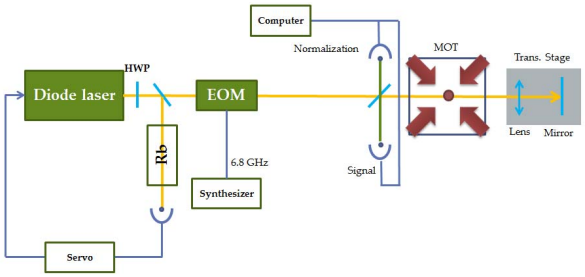


Fig. 1. Basic scheme of the cold-atom CPT clock with the EOM-based light source. The diode laser is frequency-stablized by use of the saturated absorption technique. The EOM was modulated by a 6.835 GHz microwave signal and carefully temperature-stabilized. The frequency synthesizer that drives the EOM is referenced to the 10 MHz output of a hydrogen maser so that we can perform absolute frequency measurements.

III. Effect of Improved Coherence on Light Shifts

We have performed preliminary measurements of the light shift in two EOM configurations with the lin ∥ lin CPT

Donley, Elizabeth; Ivanov, Eugene; Kitching, John; Liu, Xiaochi.
"Frequency shift mitigation in a cold-atom CPT clock."
Paper presented at 2016 IEEE International Frequency Control Symposium, New Orleans, LA, United States. May 9, 2016 - May 13, 2016.

interrogation technique and D_1 laser excitation [7]. In one configuration, we took the modulated light with all of the sidebands directly from the output of the EOM and used the entire spectrum to excite CPT signals. In a second configuration, we tried to generate a nearly perfect CPT spectrum by selecting the desired sideband by filtering the EOM output with a Fabry-Perot cavity and recombining the beam with the unmodulated laser output. In this case, the spectrum contained only the two frequency components needed for CPT excitation and no unwanted sidebands.

To characterize the light shift, we used a Ramsey period of 16 ms, a CPT pulse length of 1 ms, and we varied the intensity and optical detuning of the CPT light. We typically set the EOM modulation amplitude such that the first-order sidebands each had roughly equal power as the carrier, for which the short-term clock stability is best. We also measured the clock frequency versus sideband amplitude (discussed below).

In the simplest arrangement for interrogation with the full output spectrum of the EOM, we were able to achieve the smallest light shift. A measurement of clock frequency versus optical detuning in this configuration is shown in Fig. 2. With a high optical intensity, the frequency shift was nearly independent of intensity and detuning and was -0.002(3) Hz/MHz. This sensitivity to detuning is smaller by a factor of 350 compared to our previous system based on an OPLL [4]. At this level, the laser frequency only has to be stabilized to 340 kHz for the light shift bias to be reduced to the 1×10^{-13} level, which can be easily satisfied with the saturated absorption technique.

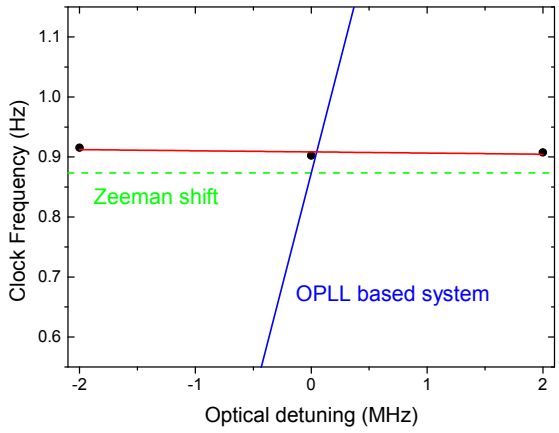


Fig. 2. Light shift vs optical detuning with 5 W/m^2 total light intensity and interrogation with all of the sidebands. The red line is a linear fit, which gives a light shift of -0.002(3) Hz per megahertz of optical detuning, which is consistent with zero. The green dashed line indicates the Zeeman shift, and the blue line indicates the light shift versus detuning observed with the OPLL-based system.

The frequency shift versus the optical detuning was also measured in the filtered Fabry-Perot-based configuration. Here, the spectrum only contained the two components needed for CPT. Although the frequency shift sensitivity to detuning was three times smaller than the previous OPLL system, it was much larger than the one shown in the Fig 2. This could be the result of a few different factors. Firstly, because of losses in the cavity, we were only able to operate the clock at intensities less than 1 W/m^2 in this configuration, at which level the coherent part of the light shift is still evident. Secondly, the commercial filter cavity that we used added significant intensity noise to the filtered component, and potentially also phase noise from the different beam paths.

IV. Frequency Shift Due to the EOM Spectrum

The measured frequency shift is about 0.06 Hz higher than the Zeeman shift (0.9 Hz) by using our new EOM-based light source (Fig. 1). The origin of this frequency shift is still to be determined, but we have found that it is sensitive to the sideband modulation index. Fig. 3 shows the frequency shift as a function of the RF power applied on the EOM.

The modulation index for EOM with a constant RF power input is known to drift due to temperature variations, photorefractive effects and aging, which could result in frequency shifts if the modulation index is not actively stabilized.

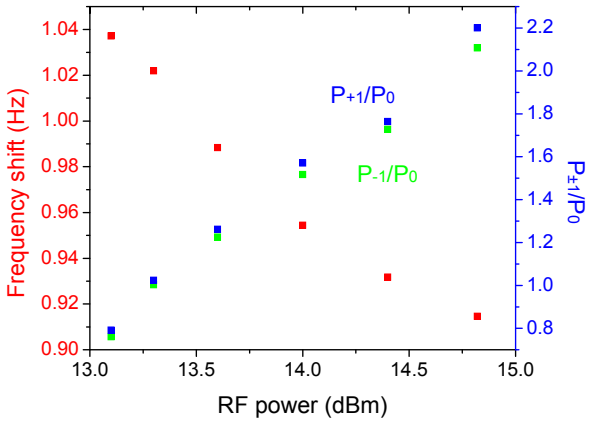


Fig. 3. Frequency shift as a function of the RF power applied on the EOM (red) and the ratio between the power of the first order sidebands and the carrier signal (blue and green). Here, 0.9 Hz is the known Zeeman shift resulting from the applied magnetic field.

V. Doppler Shift Mitigation Optical Setup

The Doppler shift is another dominant systematic that can degrade the long frequency stability of the clock. The Doppler shift introduced in our clock mainly arises because the atoms are free to move through the phase fronts of the interrogation field during interrogation and they move by more than a millimeter during the Ramsey period due to gravity.

The Doppler shift can be well cancelled in an optical set-up with counter-propagating CPT beams if the intensities of the forward- and backward-travelling beams are equal and the beams are well overlapped [3].

Donley, Elizabeth; Ivanov, Eugene; Kitching, John; Liu, Xiaochi.
"Frequency shift mitigation in a cold-atom CPT clock."
Paper presented at 2016 IEEE International Frequency Control Symposium, New Orleans, LA, United States. May 9, 2016 - May 13, 2016.

In a counter-propagating optical setup with a simple flat mirror retroreflector, the optical power of the incident beam and the reflected beam are not exactly equal because of losses from the MOT cell walls and absorption of the atoms. For our AR-coated cell, the losses from the cell walls total <2 %. Losses from atom cold-atom absorption can be much larger and can approach 10 % over the extent of the atom cloud, leaving a shadow in the retroreflected beam. To reduce the effect of the intensity imbalance resulting from the atom shadow, we use a cat's-eye retroreflector comprised of a lens and a flat mirror, with the lens positioned in front of the mirror by the lens focal length (see Fig. 4a). We position the atoms in the CPT beams such that they are in the upper half of the beams for the first Ramsey pulse and in the lower half of the beam for the second Ramsey pulse. With this approach, the Doppler shift is reduced by about one order of magnitude compared to our previous optical setup based on a simple flat retroreflector, as measured by translating the retroreflector.

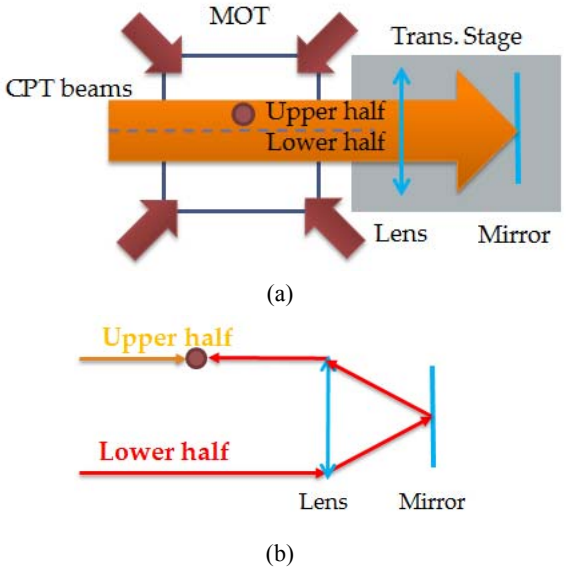


Fig. 4. a.) Doppler shift mititgation optical setup. b.) Ray traces showing the input and retroreflected rays that strike the atoms during the first Ramsey pulse. This geometry prevents the shadow from hitting the atoms for the reflected beam. For the second Ramsey pulse, the atoms are in the bottom of the beam (not shown).

With the new improvements to the light shift and Doppler shift, the short-term frequency stability of the clock is measured to be $1.5 \times 10^{-11}/\sqrt{\tau}$ and decreases after 4000 seconds to the level of 1.25×10^{-13} (Fig. 5).

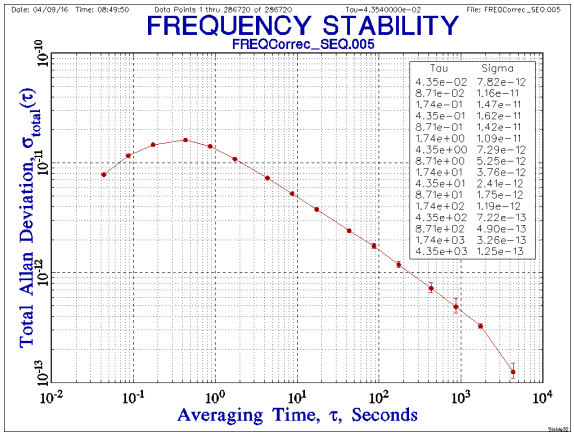


Fig. 5. Clock frequency stability.

VI. Outlook

The reduced light shift with improved CPT laser coherence is consistent with our predictions in [4]. The clock instability that we currently achieve has been improved compared to what we measured in the OPLL-based system ($4 \times 10^{-11}/\sqrt{\tau}$), but has yet to be fully optimized. Our next work will focus on stabilization of the EOM, reduction of the frequency shift introduced by the EOM, and optimizing the operating parameters.

The EOM-based system has the advantage over the OPLL-based approach in that only one interrogation laser is required. If the clock were based on D_2 interrogation, then it could be possible to perform all of the laser cooling and interrogation with a single D_2 laser and EOM modulation.

Acknowledgement

This work is funded by NIST, which is an agency of the U.S. government, and is not subject to copyright.

References

[1] J. Vanier, "Atomic clocks based on coherent population trapping: a review," *Applied Physics B,* vol. 81, pp. 421-442, 2005.

[2] E. N. Ivanov, F. X. Esnault, and E. A. Donley, "Offset phase locking of noisy diode lasers aided by frequency division," *Review of Scientific Instruments,* vol. 82, 2011.

[3] F. X. Esnault, E. Blanshan, E. N. Ivanov, R. E. Scholten, J. Kitching, and E. A. Donley, "Cold-atom double-Λ coherent population trapping clock," *Physical Review A,* vol. 88, 2013.

[4] E. Blanshan, S. M. Rochester, E. A. Donley, and J. Kitching, "Light shifts in a pulsed cold-atom coherent-population-trapping clock," *Physical Review A,* vol. 91, p. 041401, 2015.

[5] J. Ye and J. L. Hall, "Optical phase locking in the microradian domain: potential applications to NASA spaceborne optical measurements," *Optics Letters,* vol. 24, pp. 1838-1840, 1999.

[6] M. Ohtsu, *Highly coherent semiconductor lasers*. Boston, London: Artech House Publishers, 1992.

[7] A. V. Taichenachev, V. I. Yudin, V. L. Velichansky, and S. A. Zibrov, "On the unique possibility of significantly increasing the contrast of dark resonances on the D1 line of ^{87}Rb," JETP Lett., vol. 82, pp. 398-403, 2005.

Donley, Elizabeth; Ivanov, Eugene; Kitching, John; Liu, Xiaochi.
"Frequency shift mitigation in a cold-atom CPT clock."
Paper presented at 2016 IEEE International Frequency Control Symposium, New Orleans, LA, United States. May 9, 2016 - May 13, 2016.

A telecom-band cavity-enhanced single-photon source with high Klyshko efficiencies

Xiyuan Lu,[1] Steven Rogers,[1] Thomas Gerrits,[2] Wei C. Jiang,[1] Sae Woo Nam,[2] and Qiang Lin[1,*]

1. University of Rochester, Rochester, NY 14627, USA
2. National Institute of Standards and Technology, 325 Broadway, Boulder, CO, 80305, USA
**qiang.lin@rochester.edu*

Abstract: We develop an on-chip telecom-band single-photon source with Klyshko efficiencies up to 48%, the highest value for cavity-enhanced photon sources. For the first time, we relate Klyshko efficiency to high-order correlations and verify this relation experimentally.

Heralded single-photon sources are mainly based upon spontaneous parametric down conversion (SPDC) or spontaneous four-wave mixing (SFWM) in nonlinear bulk crystals, optical fibers, on-chip waveguides, or microcavities [1]. Heralding efficiency is an important parameter for applications including quantum teleportation [2] and quantum computing [3]. For photon sources based on bulk crystals, the raw heralding efficiency, i.e., the Klyshko efficiency [4], has been pushed above 82.8% at a wavelength of 810 nm [5], and 60% at a telecom wavelength [6]. Recently, microcavity-based photon sources have attracted significant interest due to their small footprints, below-milliwatt pump powers, and narrow photon spectra. However, the Klyshko efficiencies of the microcavity-based photon sources are typically limited below 10%. Moreover, no quantitative experiments have been conducted to study the heralding efficiency systematically in the photon sources with cavity enhancement.

High-Q silicon microresonator is a great photon-pair source with high spectral brightness, large coincidence to accidental ratios (CARs), and single-mode properties [7–9]. In this paper, we herald single photons out of the photon pairs generated by spontaneous four-wave mixing in a high-Q silicon microresonator. We investigate the heralding efficiency and demonstrate a Klyshko efficiency of 48%, the highest value for cavity-enhanced photon sources. Further, we find the relation between the Klyshko efficiency and high-order correlations for the first time.

Our tapered-fiber coupled photon source has the advantage of directly coupling photons from device into the single-mode fiber (Fig. 1(a)). To characterize the performance of photon sources, preparation efficiency [10] is defined as the heralding efficiency in the single-mode fiber, excluding the effects of optical components and detector efficiency. The preparation efficiency is determined by two factors, the photon coupling efficiency and the tapered-fiber efficiency. First, the photon lost inside the cavity can not be coupled out. This photon coupling efficiency η_c is related to the cavity transmissions as $\eta_c = (1 \pm \sqrt{T_c})/2$, where T_c stands for the transmissivity at the center wavelength of the cavity mode,

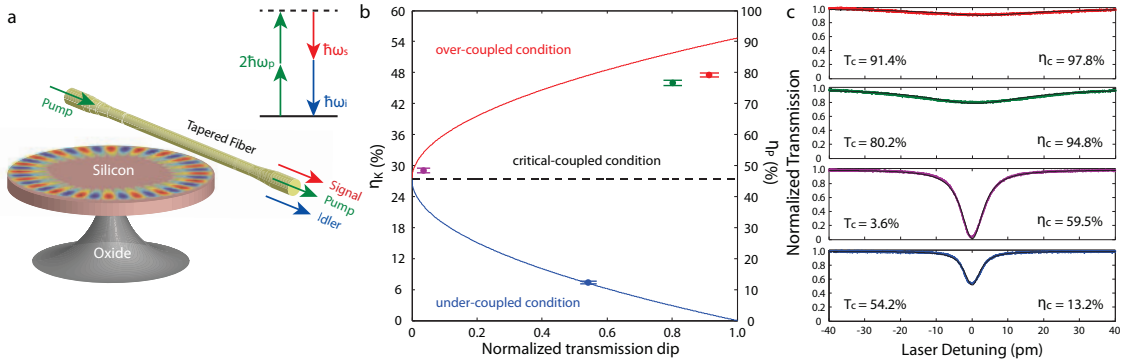


Fig. 1. (a) Schematic illustration of the silicon microresonator photon source based on spontaneous four-wave mixing, with the inset showing the energy diagram. The fiber-device distance can be tuned to change the photon coupling rate. (b) Theoretical predictions of η_K for over-coupled (red) and under-coupled (blue) conditions. Experimental data points are from various fiber-taper coupling conditions with transmission traces shown in (b). (c) The normalized cavity transmission traces with theoretical fitting in black. T_c represents the transmissivity at the center wavelength of the cavity resonance. η_c stands for the photon coupling efficiency.

Gerrits, Thomas; Lin, Qiang; Nam, Sae Woo; Rogers, Steven; Wiang, W; Xiyuan, Lu.
"A telecom-band cavity-enhanced single-photon source with high Klyshko efficiencies."
Paper presented at CLEO 2016, San Jose, CA, United States. June 5, 2016 - June 10, 2016.

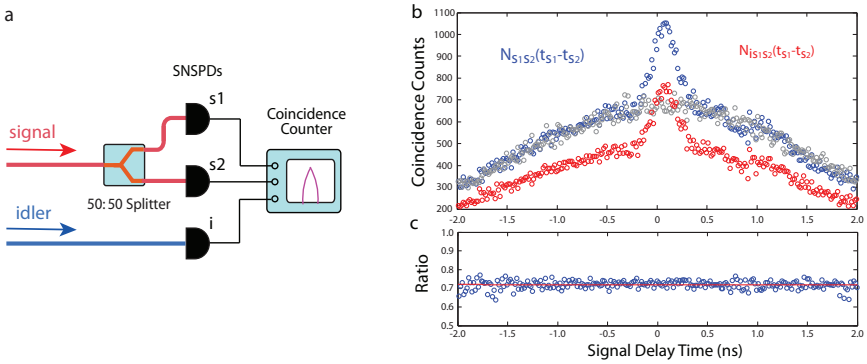


Fig. 2. (a) Scheme of the three-detector setup. (b) Coincidence counts of self correlation, background of self correlation, and triple coincidence shown in blue, grey, and red, respectively. (c) Ratio of triple-coincidence and self-correlation counts.

and the plus/minus signs correspond to the over-/under-coupled conditions, respectively (Fig. 1(c)). The red curve in Fig. 1(c) shows that the photon coupling efficiency is 97.8%, which means the photon loss inside the cavity is less than 3%. Second, the photons coupled out from the cavity can be lost in the tapered fiber. The loss of the tapered fiber before the cavity does not affect the heralding efficiency, because the photons are generated inside the cavity. Thus the efficiency related to tapered fiber is $\eta_{tf} = \sqrt{T_{tf}}$, where T_{tf} is the fiber taper transmittance. The fiber taper used in the experiment has a transmittance of 83.0%, which corresponds to a η_{tf} of 91.1%. The theoretical predictions of Klyshko/preparation efficiencies, $\eta_P = \eta_c\eta_{tf}$, are plotted in Fig. 1(b). The experimental data points are 79%, 77%, 48%, and 12%, which largely agree with our theoretical predictions. Fig. 1(c) shows the normalized transmission traces of experimental data in Fig. 1(b), with colors labeling respective data points. The over-coupled condition in red leads to a preparation efficiency of 79% and a Klyshko efficiency of 48%, among the highest values for telecom-band photon sources. The preparation efficiency (η_P) is related to the Klyshko efficiency (η_K) by $\eta_P/\eta_K = 60\%$ in this experiment, which takes into account the detection efficiency (78%) and the transmittance in optical components (77.3%).

We characterize the heralded single photons in a three-detector setup (Fig. 2(a)), which can measure triple-coincidence counts ($N_{is_1s_2}$) and self-correlation counts ($N_{s_1s_2}$) at the same time. We find that Klyshko efficiency can be verified by the ratio of these counts ($\eta_R \equiv N_{is_1s_2}/N_{s_1s_2}$) as $1-\eta_R = (1-\eta_K)^2$, because the only cases that no idler photons are heralded in triple coincidence are those in which both signal photons fail to herald the idler photons. To verify this relation, we compare the triple-coincidence and self-correlation counts in Fig. 2(b). The ratio of triple coincidence counts to self correlation counts (η_R) is plotted in Fig. 2(c), which is a constant of 72% regarding signal delay time. The Klyshko efficiency is thus calculated to be 47%, which is very close to the experimental value of 46% (the green dot in Fig. 1(b)). This is the first time that high-order correlations are used to verify Klyshko efficiency.

In sum, we develop a telecom-band heralded single-photon source in a high-Q silicon microresonator. We investigate its Klyshko/preparation efficiency and demonstrate a Klyshko efficiency of 48%, the highest value for cavity-enhanced photon sources. We also demonstrate a method to verify Klyshko efficiency by high-order correlations.

1. B. Lounis and M. Orrit, *"Single-photon sources,"* Rep. Prog. Phys. **68,** 1129-1179 (2005).
2. B. J. Metcalf et al., *"Quantum teleportation on a photonic chip,"* Nature Photon. **8,** 770-774 (2014).
3. P. Kok et al., *"Linear optical quantum computing with photonic qubits,"* Rev. Mod. Phys. **79,** 135-174 (2007).
4. D. N. Klyshko, *"Use of two-photon light for absolute calibration of photoelectric detectors,"* Sov. J. Quantum Electron. **10,** 1112 (1980).
5. M. Giustina et al., *"Bell violation using entangled photons without the fair-sampling assumption,"* Nature **497,** 227-230 (2013).
6. S. Wollmann, M. Weston, H. Chrzanowski, and G. J. Pryde, *"High heralding efficiency single photon source at telecom wavelength,"* CLEO/Europe-EQEC, Paper EB-3.2-wed (2015).
7. W. C. Jiang, X. Lu, J. Zhang, O. Painter, and Q. Lin, *"Silicon-chip source of bright photon pairs,"* Opt. Express **23,** 20884-20904 (2015).
8. X. Lu, W. Jiang, J. Zhang, and Q. Lin, *"High-purity single-mode photon source for integrated quantum photonics,"* SPIE Sensing Technology and Applications, 950018-9, Baltimore, MD (2015).
9. S. Rogers, X. Lu, W. C. Jiang, and Q. Lin, *"Twin photon pairs in a high-Q silicon microresonator,"* Appl. Phys. Lett. **107,** 041102 (2015).
10. J. B. Spring et al., *"On-chip low loss heralded source of pure single photons,"* Opt. Express **21,** 13522-13532 (2013).

Gerrits, Thomas; Lin, Qiang; Nam, Sae Woo; Rogers, Steven; Wiang, W; Xiyuan, Lu.
"A telecom-band cavity-enhanced single-photon source with high Klyshko efficiencies."
Paper presented at CLEO 2016, San Jose, CA, United States. June 5, 2016 - June 10, 2016.

Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results

Qian Gong[a]*, David Content[a], Margaret Dominguez[a], Thomas Emmett[a], Ulf Griesmann[b], John Hagopian[c], Jeffrey Kruk[a], Catherine Marx[a], Bert Pasquale[a], Thomas Wallace[a], Arthur Whipple[d]

[a]NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD, USA 20771
[b]National Institute of Standards and Technology, Gaithersburg, MD, USA 20899
[c]Lambda Consulting, Harwood, MD USA 20776
[d]Conceptual Analytics, LLC, Glenn Dale, MD, USA 20769

ABSTRACT

The slitless spectrometer plays an important role in the WFIRST mission for the survey of emission-line galaxies. This will be an unprecedented very wide field, HST quality 3D survey of emission line galaxies[1]. The concept of the compound grism as a slitless spectrometer has been presented previously. The presentation briefly discusses the challenges and solutions of the optical design, and recent specification updates, as well as a brief comparison between the prototype and the latest design. However, the emphasis of this paper is the progress of the grism prototype: the fabrication and test of the complicated diffractive optical elements and powered prism, as well as grism assembly alignment and testing. Especially how to use different tools and methods, such as IR phase shift and wavelength shift interferometry, to complete the element and assembly tests. The paper also presents very encouraging results from recent element tests to assembly tests. Finally we briefly touch the path forward plan to test the spectral characteristic, such as spectral resolution and response.

Keywords: grism, slitless spectrometer, high efficiency diffractive surface

1. INTRODUCTION

The Wide-Field Infrared Survey Telescope (WFIRST) is a NASA observatory designed to perform wide-field imaging and slitless spectroscopic surveys of the near infrared (NIR) sky for the community[2]. The current WFIRST design of the mission makes use of an existing 2.4m diameter telescope to enhance sensitivity and imaging performance. The WFIRST payload includes two main instruments: a wide field instrument and a coronagraph instrument. The wide field instrument provides the wide-field imaging and slitless spectroscopy capability required to perform the dark energy, exoplanet microlensing, and NIR surveys. A compound grism design is selected as the slitless spectroscopy. More specifically, the slitless spectroscopy is going to survey the emission-line galaxies to cover red-shift $1 < z < 6$ (Figure 1). However, the emphasis of this paper is to address how to design and build the slitless spectrometer as the part of the risk reduction of WFIRST pre-phase A study.

What distinguishes WFIRST from other previous and existing space telescopes is its large FOV, higher spectral resolution, and a faster beam. All these bring new challenges to grism optical design and fabrication. The detailed comparison, challenges, solutions, and prototype development are the discussion through this paper.

*qian.gong-1@nasa.gov; phone 1 301 286-1490; fax 1 301 286-7230; nasa.gov

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

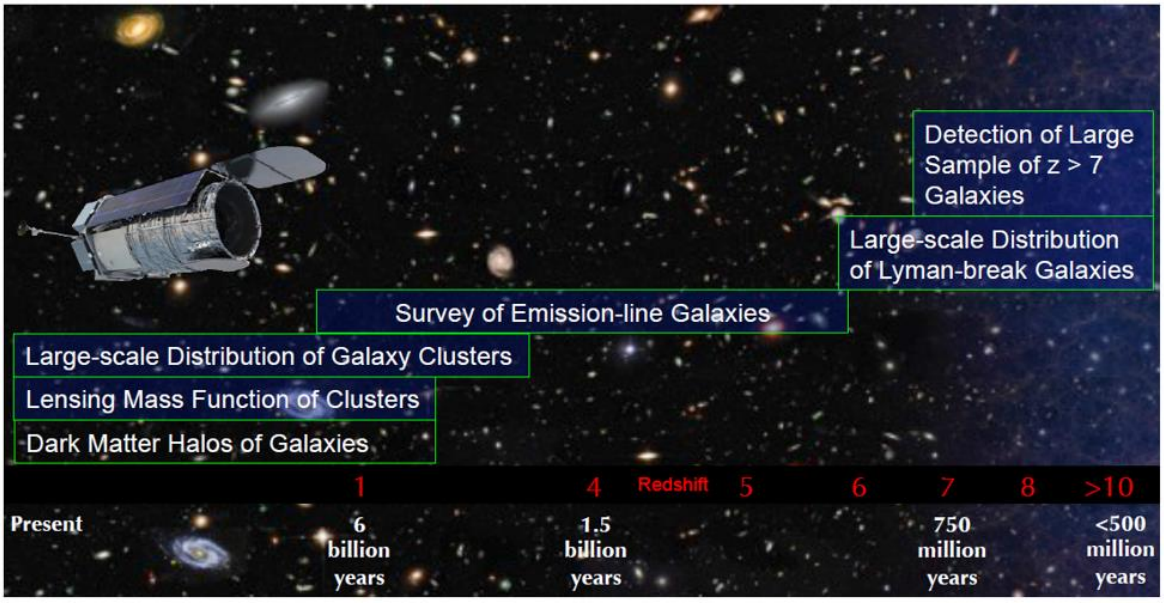


Figure 1. WFIRST-AFTA: A unique probe of cosmic structure formation history.

2. GRISM SPECIFICATION AND CHALLENGES

2.1 History of Compound Grism Optical Design

The WFIRST pre-phase A has gone through 6 study cycles. The first cycle of the compound grism design started in 2013. In the first 3 cycles, a four element compound grism was the baseline design. Almost all optical surfaces were aspheric, and the diffractive surface, grating, was on one of the aspheric surfaces. Even though selected and varying NIR glass materials were used, the image quality was still not satisfactory. So starting from cycle 4, the addition of a second diffractive surface was introduced. The aberration created by the grating in a converging f/8 beam with a large FOV was easily corrected by the diffractive corrector. This design is used for building a prototype grism assembly. A NASA New Technology Report was filed based on this design[3]. The subsequent Cycle 5 grism design does not have any major change from cycle 4. However, in the mission concept review design (MCR or "cycle 6"), the minimum wavelength has decreased from λ= 1.35 μm to 1.0 μm. Meanwhile, in order to split some visible light for grism fine guidance, a dichroic beam splitter at a fairly large incidence angle has been inserted before the grism dispersion. This dichroic becomes one of the elements in the grism assembly. These changes have made the cycle 6 grism design very different than the current prototype design. In the following sections, both prototype and cycle 6 designs are discussed and compared.

2.2 WFIRST Compound Grism Specification

The wide-field grism assembly is to be inserted into a ~f/8 beam around the pupil image plane as one element in the filter wheel of the wide-field instrument. The goal is to provide the dispersion with the spectral resolution R ~700 in the wavelength range from 1.35μm to 1.95μm for prototype, and that from 1.0μm to 1.9μm for cycle 6 study. The detailed specifications for prototype and cycle 6 study are listed in Table 1. In addition to the specification, the following restrictions have to be met simultaneously.

1. When inserting the grism into the f/8 beam, the image plane has to be unchanged in order to be parfocal (same focal position as the filters held in the same element wheel) with other filters.

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

2. The position of each ray from central wavelength has to be unchanged after the dispersion. This implies that a prism has to be used to make the compound grism assembly zero deviation.
3. Because the grism is one of the elements in the filter wheel, the available volume is very limited. The grism has to have a very compact design. For cycle 6 study, a dichroic beam splitter is inserted before the grism to reflect non-signal visible light onto 2 detector arrays for providing fine guidance during exposures. The requirements and restrictions above have made this grism design much more challenging from design to fabrication, to alignment, and to test. In section 2.3, the grism of HST wide field camera 3 will be used as an example to show why WFIRST grism is so challenging.

Table 1. Grism specification for prototype and cycle 6 design.

	Prototype	MCR design
Wavelength range (μm)	1.35 – 1.95	1.0 – 1.9
FOV (°)	0.788 x 0.516	Overall area is unchanged Shape has minor change
Average beam diameter at grism (mm)	120	120
Beam f/-ratio at grism (mm)	~f/8	~f/8
Wavefront error	Diffraction limited	Diffraction limited at $\lambda > 1.3\mu m$, Spot size similar
Spectral resolving power (per 2 pixels)	645 – 900 (461×λ)	461 – 876 (461×λ)
Capability	Dispersion only	Dispersion and pointing for spectrometer
Compactness	70mm total thickness for a fixed diameter ~120mm	w/o dichroic, total thickness is ~52mm, with dichroic ~80mm

2.3 Grism Comparison: WFIRST versus HST

Grisms have been used in a number of HST instruments, Wide Field Camera 3 (WFC3)[4], Near Infrared Camera and Multi-Object Spectrometer (NICMOS)[5], Advanced Camera for Surveys (ACS), etc. A single grism itself was used there. For WFIRST, it is totally different, controlling the image quality is greatly challenging even using compound grism assembly. The parameters in Table 2 shows the differences between WFIRST and HST.

Table 2. Optical system comparison between WFIRST and HST.

	WFIRST	**HST WFC3**
FOV	0.28 degree^2	0.0012 degree^2
Spectral resolving power	R = 700	R = 130
f/#	8	11

The 3 parameters listed in Table 2 are the key parameters in optical design to estimate how challenging the spectrometer optical design is. The WFIRST FOV is more than 100 times larger than HST WFC3. As we know, the off-axis aberration is a function of FOV, depending on the type of the off-axis aberration, the field dependent can be linear, quadratic, cubic, etc. For example, when the field angle increases 10 times, the astigmatism will increase 100 times! Next, if the grism is in collimated space, no aberration is created. When the f/# gets smaller and smaller, the aberration from the dispersive

element becomes more and more difficult to control. The f/# of WFIRST is f/8, but HST WFC3 is slower at 11. The third, as a spectrometer, the difficulty is also a function of spectral resolution. The higher of the resolution, the more difficult it is. Besides all this, the aberration is also a function of the aperture. The larger the aperture, the more difficult it is, even with the same f/#. The size of the WFIRST grism (~120mm in diameter) is much larger than the grism on WFC3.

3. OPTICAL DESIGN OF GRISM ASSEMBLY

3.1 Optical Design of Prototype

As mentioned above, the optical design of the prototype uses two diffractive surfaces. The reason for adding one more diffractive surface is that a grating in a non-collimated space introduces a lot of aberration. This aberration is not only a function of incident angle, but also linear to the wavelength. It is really difficult to compensate the aberration scaled to wavelength using the conventional optics, even when freeform aspheric surfaces are used. However, aberration created from diffractive surfaces is all scaled to wavelength. Based on this, adding another diffractive surface to compensate the aberration created by the grating becomes a natural solution for this grism. The additional diffractive surface is a very weak Fresnel lens like surface with non-symmetry in x and y directions. The off-axis aberration created by this surface efficiently compensate the aberration created by the grating surface, which makes the final image quality from design residual diffraction limited. With the second diffractive surface, the entire optical design has simplified. All element material is fused silica. The surfaces are either flat or spherical. The whole assembly is more compact to fit into the very limited space. Figure 2 shows the prototype grism optical design and the grism position in the WFIRST wide-field instrument.

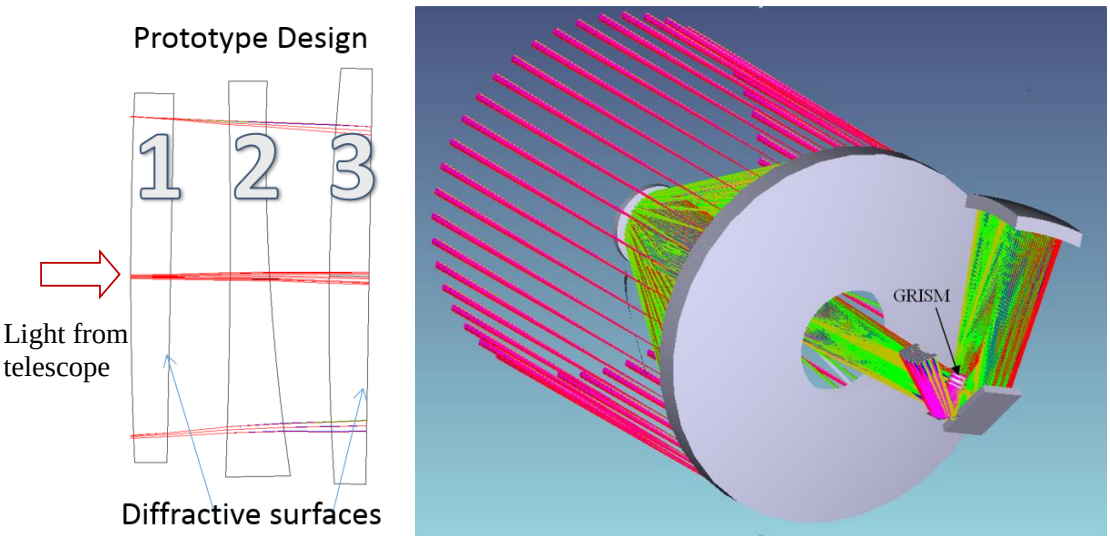


Figure 2. On the left is the prototype grism layout. On the right is the grism with the WFIRST wide-field instrument.

The main function of each element is as following: 1. Element #1 is an aberration corrector. It is to compensate the wavelength scaled aberration created by the grating (Element #3). 2. Element #2 is a prism. It is to make the whole grism assembly zero deviation at the central wavelength. Both surfaces of the prism are spheres for aberration control. 3. Element 3 is the key element of the grism, it provides the required spectral dispersion (R = 461λ). The grating in this assembly is not a regular straight line grating. It has curved lines and variable line spacing for further aberration control. The patterns of the two diffractive surfaces, the corrector and the grating, are shown in Figure 3 (a) and (b). The optimized grism design provides a diffraction limited performance for all wavelength and fields. Figure 4 shows the spot diagram of the grism design across the FOV and wavelength range meets the specification.

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

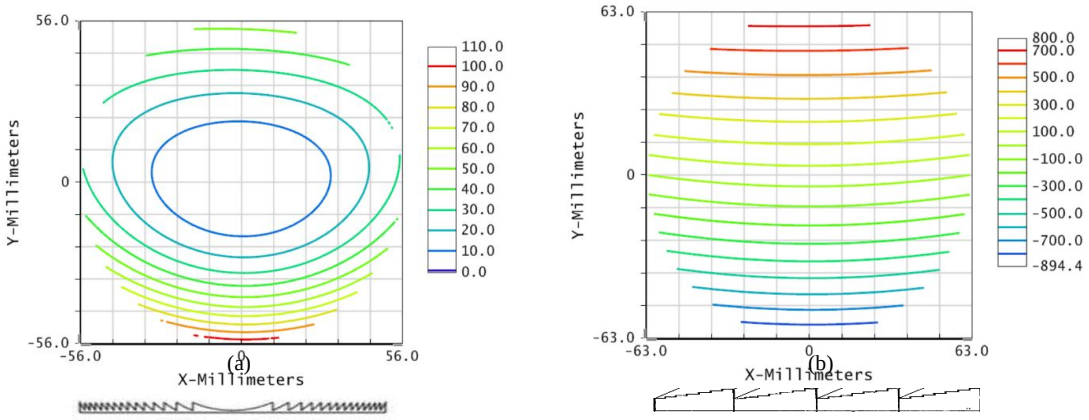


Figure 3. Diffraction patterns of Elements #1 and #3 diffractive surfaces. (a) is for Element #1 which is made by grey scale process and (b) is for Element #2 by multi-step overlay process.

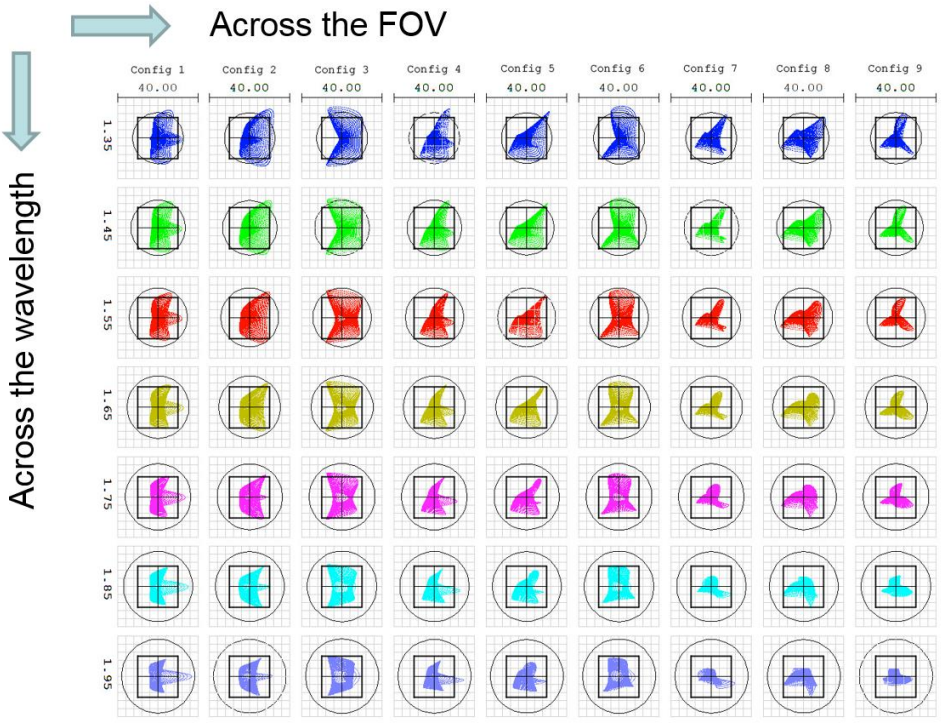


Figure 4. Diffraction limited performance for the entire FOV and entire wavelength range.

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

3.2 Optical Design of Cycle 6

The cycle 6 optical design is similar to the prototype in principle. There are two major differences. The first is the wider wavelength range. In cycle 6 the wavelength range has extended to 1.0 μm to 1.9 μm compared to 1.35 μm to 1.95 μm in prototype. The second one is that a dichroic beam splitter is added before the grism to provide a non-dispersive image for fine guidance sensor. This dichroic has to be tilted enough to avoid the reflected beam from interfering with the incoming beam and the structures in the payload. As a result, the rest of the grism elements have to be tilted too to minimize the aberration. Figure 5 shows the cycle 6 optical design and its dual functions in the WFIRST instrument.

The function of the three grism elements are similar to the prototype design. The added dichroic beam splitter reflects the visible light onto the silicon detector array to provide fine guidance for the grism assembly. The image quality of this design is similar to the cycle 5 design measured by spot size. But the wavelength range shifts to shorter wavelength, and the diffraction limited performance is scaled to wavelength, so it is not completed diffraction limited for the entire wavelength range.

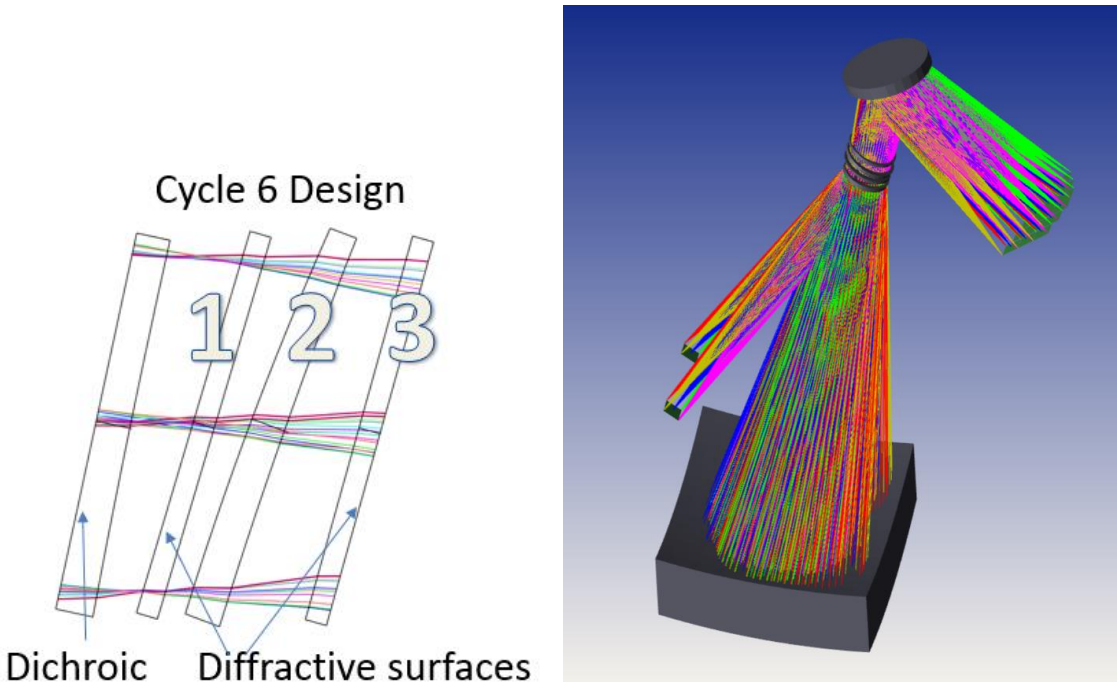


Figure 5. Cycle 6 grism optical design. On the left is the grism layout. On the right is the grism with the beam splitter to direct the visible light to guidance detectors. Ray trace from M3 is not shown for visibility.

4. CHALLENGES AND SOLUTION FOR PROTOTYPE

There are many challenges in the process of building the prototype. In this section, we discuss each challenge and how the challenge was solved.

4.1 Diffraction efficiency of the DOE (Diffractive Optical Element) surfaces

The two DOE surfaces in the design greatly improve the spectrometer performance. However, it requires high diffraction efficiency for each DOE surface in order to obtain the high throughput and reduce the ghosts from unwanted orders. To solve this problem, microlithography and reactive plasma etch (dry etch) technology was selected to make the DOE surfaces. In the optical design phase, this problem had already been considered. So fused silica material was

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

selected for the elements with DOE surfaces. Fused silica is a dry-etch friendly material. The MEMS companies have accumulated extensive experience for dry-etching the patterns into the fused silica substrate with high efficiency close to 100% at the selected peak wavelength, which reaches theoretical prediction[6].

It is worth to mention that we have tried two different dry etch technologies: gray scale and multi-step overlay. For the low density frequency, < 5 grooves / mm, either technology works. However, for the higher density, multi-step overlay yields much higher diffraction efficiency. This is because the gray scale is not able to get very sharp tops and bottoms. This is also because the slope of the gray scale is not as steep as multi-step overlay. This is also why the mechanically ruled gratings can only provide a peak efficiency of about 80%.

4.2 Dry-etch diffractive pattern onto a substrate with a wedged convex back surface

Most of the dry etch equipment is made for semi-conductor industry to etch flat wafers. Our substrates are not only too thick (as shown in Figure 2) but also wedged with a curved back surface. We have tried two ways to solve this problem. For element #1 with a coarse groove density, we found a vendor having a gray scale lithography equipment that can accommodate our thick substrate. However, a special custom mount had to be made to hold the curved back for laser writing and etching. For element #3, the grooves are too dense for gray scale. Multi-step overlay was selected. But the equipment can only accept flat wafers less than 6mm thick. So we developed an innovative way to handle it. We made two separate parts: a thin flat substrate and a thicker substrate with a convex surface. The total thickness is the same as the specified thickness for Element #3. We first etched the diffractive pattern into the thin flat substrate, then use adhesive free bonding technology to bond them together. Elements #1 and #3 have been successfully been fabricated, delivered and tested at NASA Goddard.

4.3 Complex element tests

The element tests are also challenging. The element #2 is a powered prism, which is the easiest of all three. Because it does not have a diffractive surface, so it can be tested at any wavelength. However, the two concave surfaces on the prism also introduce enough aberration to make the test challenge for normal interferometer test. The solution is to design a special Computer Generated Hologram (CGH)[7] to compensate the huge aberration when testing the powered prism using a Zygo interferometer at HeNe wavelength.

The two other elements, elements #1 and #3, are even more challenging due to the diffractive surface on one surface. It is known that the aberration introduced by a diffractive surface has a strong wavelength dependency, usually proportional to wavelength. Because the operation wavelength range is from 1.35 μm to 1.95 μm, the CGH has to be designed in that wavelength range, and an IR interferometer has to be used to test them. In the relatively few IR interferometers that work in our wavelength range, λ= 1.55μm is popular. So the CGHs were designed for that wavelength. It is probably worth to mention that we also tried to design a CGH to test the elements at HeNe wavelength. The design itself works. However, the element's diffractive surfaces are designed and fabricated to provide highest diffraction efficiency at peak wavelength of 1.65 μm. The diffraction efficiency at 0.6328 μm is too low to be seen. During the test, the designed first order signal is buried in ghost fringe. We have successfully designed IR CGHs and tested the elements using IR interferometer. The details will be described in the latter part of the paper.

4.4 Mechanical mount design to minimize wavefront distortion at cryogenic temperature

Based on the operation wavelength range and the signal to noise analysis, the grism needs to operate at temperatures from 150K to 170K. Therefore, the mount design has to take the temperature change into account. Two approaches are used for the design: 1. The mount material is selected as Invar to best match the fused silica CTE (Coefficient of Thermal Expansion). 2. A special feature to the ring was designed to further reduce the stress at cryo. Besides the operation at cryo, the very limited volume is another challenge to the mechanical assembly design. As for solutions, low CTE Invar was selected as the mount material. Very compact mounts and assembly with special wire-EDM features are built-in to control the CTE introduced stress on elements.

4.5 Alignment and assembly challenge

The grism optical alignment presents another challenge to the prototype. Each element has only one axis of symmetry, which is already a challenge to the alignment. The diffractive surfaces make the process far more challenging. For this

prototype, we tested the assembly based on the mechanical tolerance and fine tuning under the interferometer to see if the complicated optical alignment can be eliminated. This method was successful.

5. GRISM OPTICAL ELEMENT TEST

All optical elements have been completed and delivered. They are undergoing a series of tests and characterizations. The warm test is completed. The cryo test plan and setup has been prepared. However, we are still looking for a vibration insensitive IR interferometer to overcome Dewar vibration.

5.1 Grating coupon diffraction efficient test

In order to assure that the grating efficiency meets our prediction and support the WFIRST throughput requirement, grating samples have been designed, ordered, delivered, and characterized. A sample blazed grating using multiple overlay technology has been fabricated by JenOptik Inc. A visible grating was initially designed to allow us to use the available sources and the detector. For the NIR wavelength of the grism in 1.35 µm – 1.95 µm, the period is greater, the fabrication difficulty and scattering effect are reduced. Therefore if the high efficiency in the visible can be reached, it will not be a problem for longer wavelengths. Figure 6 shows the diffraction efficiency measurement setup.

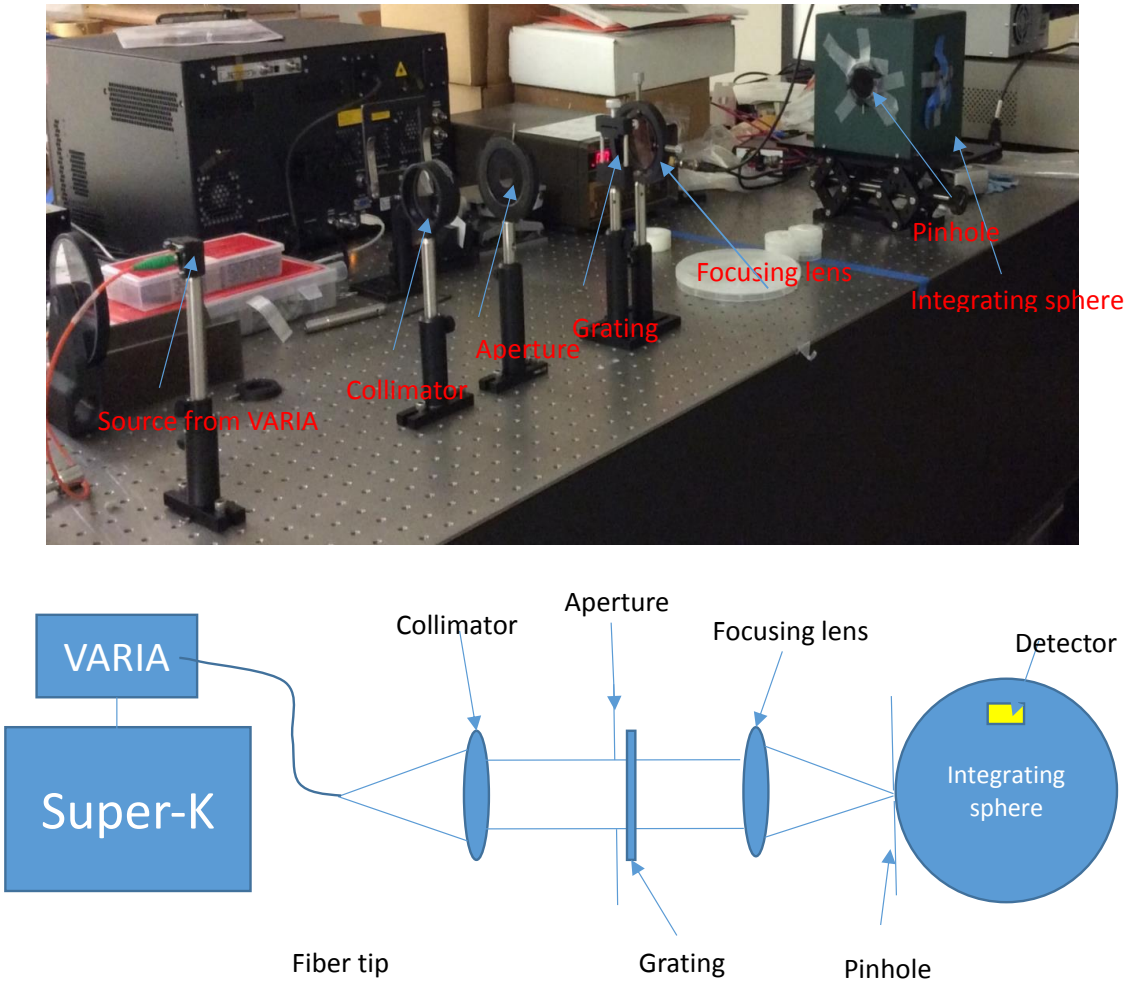


Figure 6. The photo on top is the diffraction efficiency measurement setup, At bottom is a schematic of the measurement.

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

A super-continuum source from NTK was selected as the light source. The light from the source is directed into VARIA, a tunable single line filter. The output of the VARIA goes to a single mode fiber with a wavelength bandwidth of 4nm. The tip of the fiber is positioned at the focus of a collimator. The grating is in the collimated space. An aperture is inserted in front of the grating to make sure all light go through the grating area. A focusing lens is positioned after the grating to focus the light into a pinhole at a port of an integrating sphere. The pinhole size is adjusted to pass only designed diffraction order. During the test, we adjust the central wavelength and record the output of the desired order. Then we remove the grating and let all light received by the integrating sphere. The diffraction efficiency is the ratio of the throughput from the desired grating order to the total throughput without the grating. Figure 7 shows the measured diffraction efficiency versus wavelength.

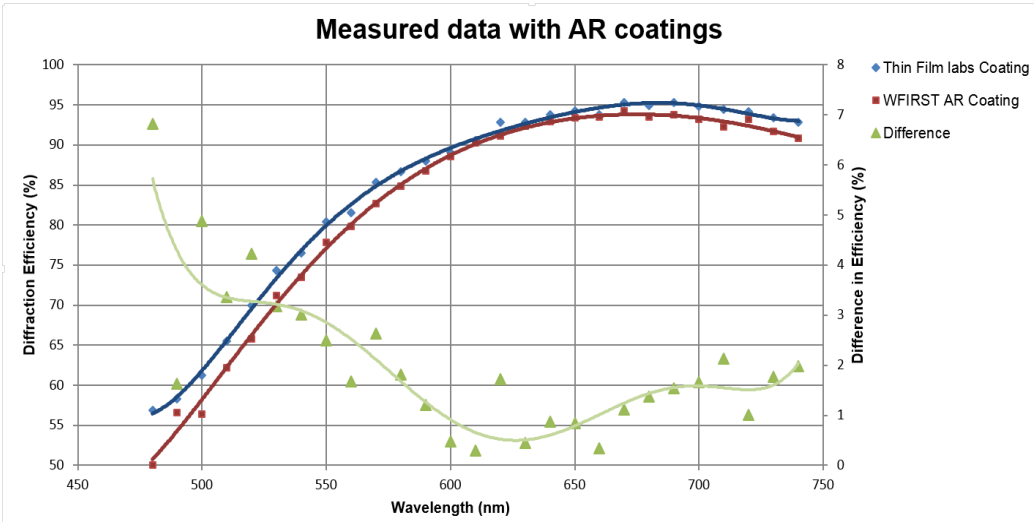


Figure 7. Measured diffraction efficiency of the visible wavelength initial grating coupon.

The measurement wavelength starts at 480nm, the shortest wavelength the super-continuum source provides, and ends at 740nm. This range contains the wavelength with the design peak efficiency. The shape of the curve fits very well with the theoretical prediction[6]. The blue and red curves are measured for two different antireflection coatings. The green curve shows the difference of the two coating using the scale on the right.

Figure 8 shows a theoretical prediction of diffraction versus wavelength in the defined 1.35 μm – 1.95 μm range. It is noted that the efficiency drops by about 10% for the two edge wavelengths.

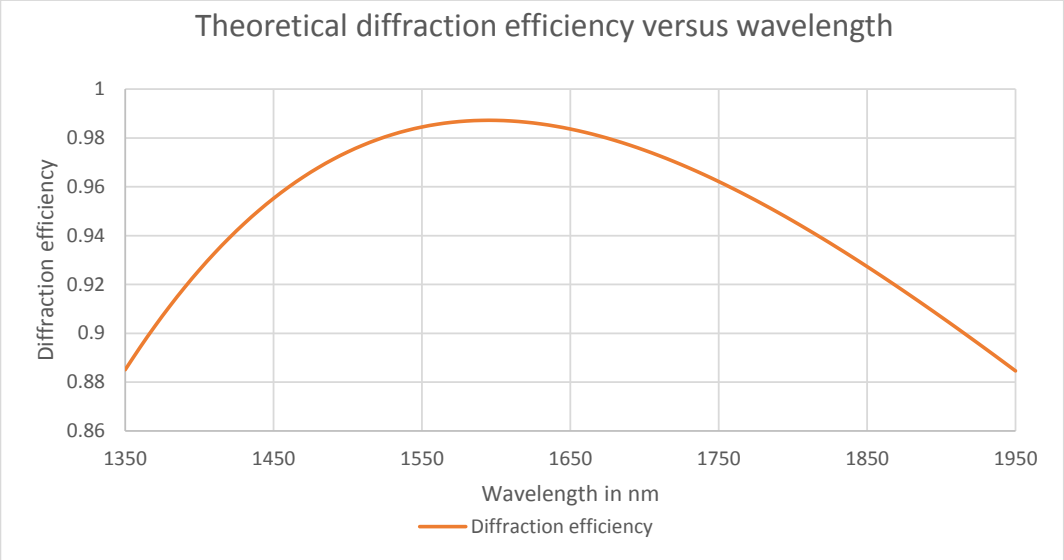
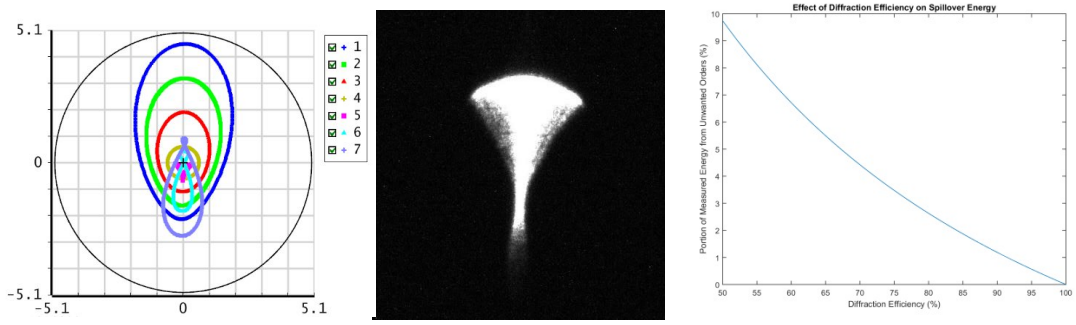


Figure 8. Theoretical diffraction efficiency prediction based on 16 step overlay.

5.2 Diffractive Corrector Efficiency Measurement

The aberration corrector is a slow freeform Fresnel lens. The light from different orders can't be completely separated. This makes the efficiency measurement very difficult. Plus, the wedged convex back surface adds more aberration to make the spot from the wanted order very large. Before the measurement, we modeled it using Zemax. The result shows that even if the diffraction efficiency is not very high, the measurement still holds, by calculating spillover power in the area that other orders overlap the wanted order.

Figure 9 (a) shows the computer generated spot diagram with different orders. (b) is the image of the real measurement at 1590nm, and (c) is the curve indicates the measurement error with respect to element diffraction efficiency. The measurement setup is very simple. The Element #1 was inserted to a collimated beam. An IR detector array was positioned to the focus of Element #1. During the measurement, a special effort was made to find unwanted orders. But they are all too weak to be found, buried under the detector array noise. This indicates that the diffraction efficiency of this diffractive aberration corrector is high, On the other hand, it implies that when working together with other grism elements, the wanted order forms a tiny spot on the final WFIRST detector array, the ghost from unwanted orders will be way below the detector noise.



(a) (b) (c)

Figure 8. (a) Spots from all diffraction orders from -3 to +3. 1st order (wanted order) is in pink. (b) Image of the only observed 1st order spot. (c) Measurement error versus diffraction efficiency.

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

5.3 Grism Elements Test

5.3.1 Element #1: Aberration corrector

As mentioned in Section 3, the element #1 is to correct the aberration created by the grating in the f/8 space. The first surface is a flat with a diffractive pattern etched into the surface. The second surface is a convex, but the axis of the vertex is tilted relative to the first surface. After a market survey, we realized that only one company provided a quotation based on best effort. Fortunately, this best effort turned out to be good enough. The measured results show that both diffraction efficiency and wavefront are good. Figure 10 shows the photo of the mounted grism elements, Element #1 is on the right.

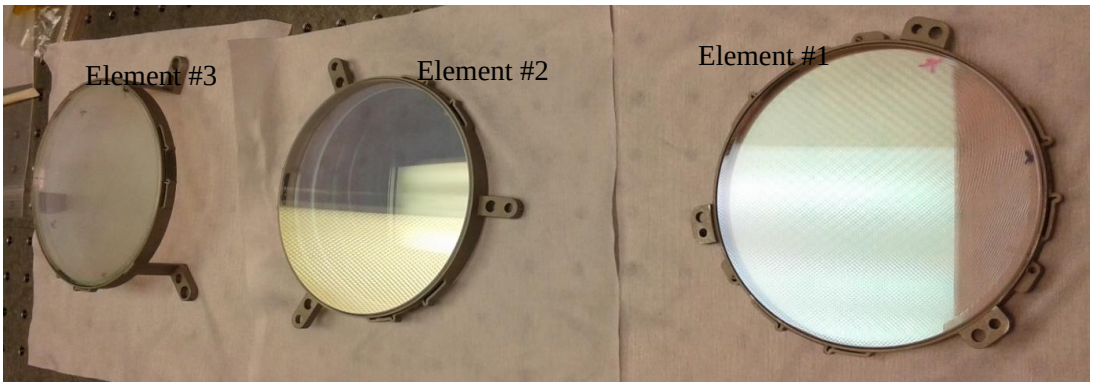


Figure 10. Photo of completed and mounted prototype grism elements #1, #2, and #3.

The wavefront measurement was designed using an interferometer with a specially designed CGH. Based on the available IR interferometer wavelength, the CGH is designed for λ= 1.55μm. We started with a visible CGH design, but the unwanted diffraction orders dominate the light returned to the interferometer. No meaningful interferogram was observed. It also needs to mention that the CGH needs to be a phase CGH in order to obtain a good fringe contrast in the interferogram, though it does not have to be blazed. Figure 11 shows the designed measurement setup. It is seen from Figure 11 that the interferometer has a collimated beam output beam. The CGH sits in the collimated space. The advantage of having the CGH in collimated space is that the CGH displacement relative to the interferometer is not sensitive. When we design the CGH, we have in mind to use the CGH for both ambient and cryo test. The elements in cryo Dewar usually move a little from ambient to cryo, so we could compensate the motion by moving the CGH accordingly. The focus is used as a carrier for this CGH design. It matches the weak power of the Element #1 perfectly, so the beam after the Element #1 is collimated again. Only a flat mirror is needed to return the beam back to the interferometer. In this case, the mirror displacement is not sensitive either, only tip/tilt angle is critical. The angular adjustment is straight forward under the guidance from interferogram fringes. Figure 12 shows the pattern of the CGH. The existing non-symmetry is to compensate the aberration introduced by the wedge of Element #1. Because of the existing central obscuration in the Element #1, the central portion of CGH is used as a metrology pattern to align the Element #1 relative to the interferometer.

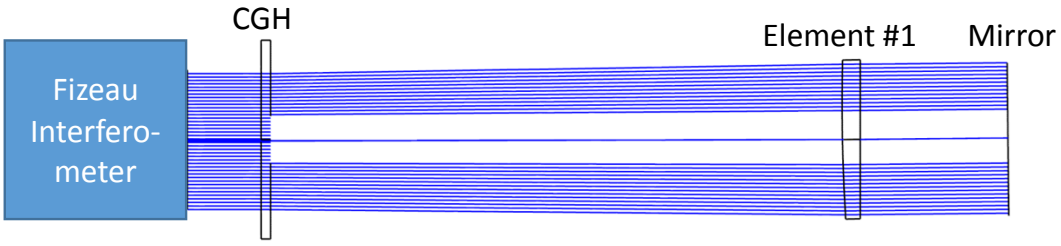


Figure 11. Layout of designed Element #1 wavefront measurement. The focus carrier of the CGH is designed to match the power of Element #1 to make the beam at the return mirror collimated.

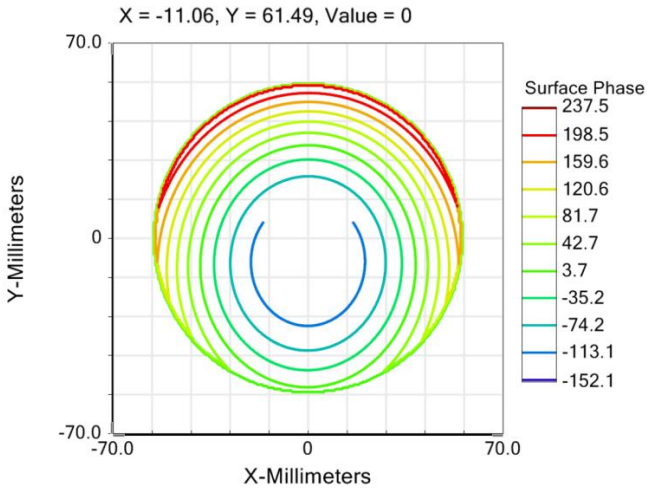


Figure 12. Hologram pattern of CGH for Element #1.

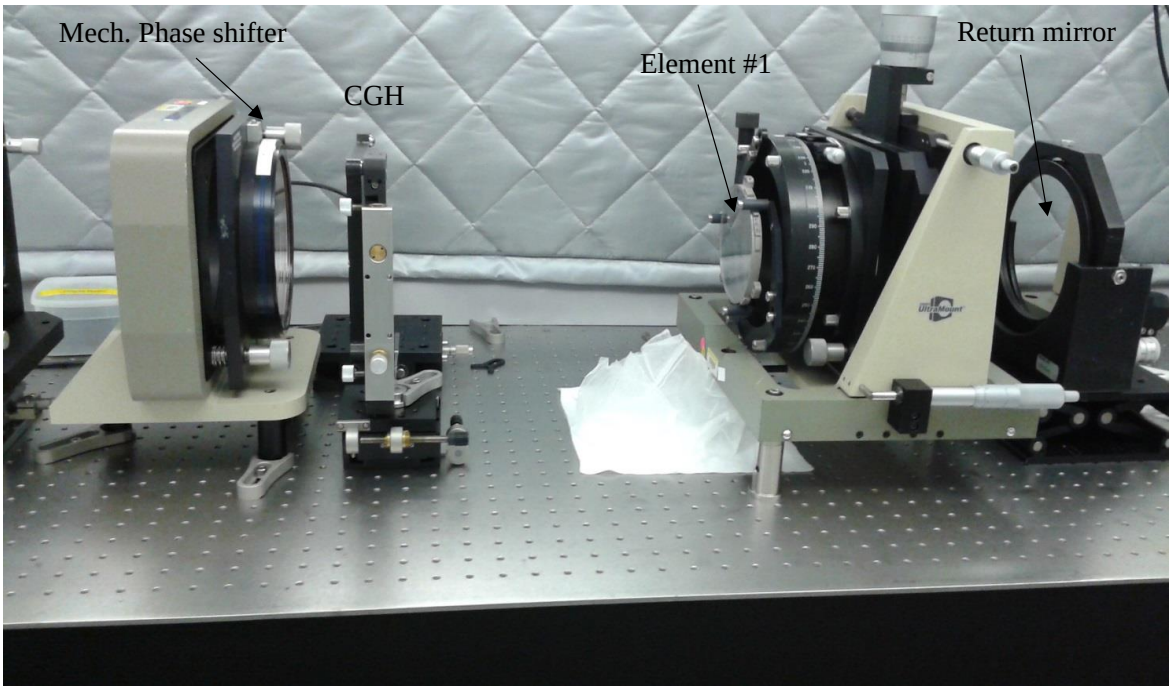


Figure 13. Element #1 test setup in the NIST IR^3 infrared interferometer. This setup was used for both mechanical phase shift and wavelength phase shift measurements. During the phase shift, the wavelength is fixed, and during the wavelength shift, the phase shifter is fixed.

The transmitted wavefront measurements for Elements #1 and #3 were made with both piezo-mechanical[8] and wavelength phase shifting interferometry to have greater confidence in the measurement results. The mesurements were made using two different interferometers, the improved infrared interferometer (IR^3) at NIST and a Zygo interferometer at NASA. The IR^3 interferometer was originally developed for thickness measurements of silicon wafers[9]. For the measurements of the WFIRST elements the interferometer was modified with a new collimator lens and a piezo-mechanical phase shifter. The

interferometer light source is a single-mode tunable diode laser with a wavelength range centered at 1550 nm. The piezo-mechanical shifter had a range of about 420° at 1550 nm, sufficient for basic phase shifting algorithms. We chose a phase shifting algorithm with 7 samples and 60° phase shift between samples for the measurements with mechanical phase shifting. For the measurements with wavelength phase shifting we used a phase shifting algorithm with 13 samples and 60° phase increment between samples. Due to the long interferometer cavity the wavelength change required to effect the 4π phase shift was very small. Based on the specifications of the diode laser and the operating parameters used we estimate that the wavelength change during a phase measurement was approximately 0.01 nm.

The collimated test beam from the interferometer is first incident on a transmission flat (with flatness error $\lambda/30$ at 633 nm) that is mounted on the mechanical phase shifter. The transmission flat can be seen on the letft in Figure 13. Because the CGH is in a collimated space, a transmission flat can be inserted with negligible additional aberration. When the beam passes the CGH, it slowly diverges to match the power of Element #1, meanwhile the same amount of aberration with opposite sign is generated to cancel the aberration of Element #1. In addition to the measurements made with the IR^3 interferometer, the same measurement was also performed using a IR Zygo interferometer at GSFC with wavelength shifting only. The tested wavefront errors from all three measurements are listed in Table 3. Figure 14 is one of the wavefront maps measured by IR^3 phase shift interferometer, but analyzed with Zygo software. The test result shows that the Element #1 meets the design requirements.

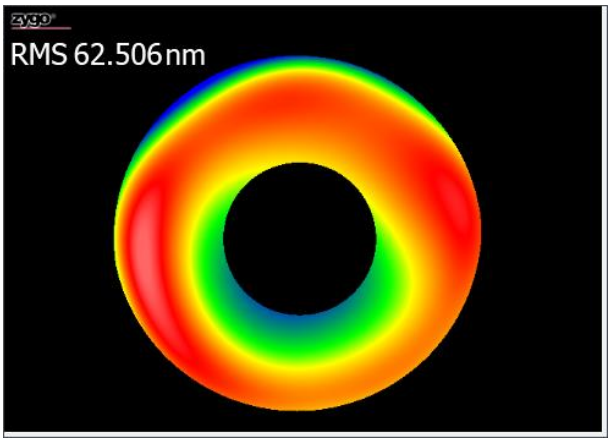


Figure 14. Zernike polynomial fit to a wavefront map of Element #1 that is measured by NIST IR^3 phase shift interferometer. The raw wavefront data is then processed using Zygo software MX .

Table 3. Measured Element #1 wavefront error using 3 different means.

	NIST IR^3 piezo shifting	NIST IR^3 wavelength shifting	Zygo wavelength shifting
Wavefront error (RMS)	63 nm	63 nm	88 nm

The data in Table 3 show that the results from the mechanical phase shifting and wavelength shifting at NIST are identical. This demonstates that the wavelength shift for our long optical path length works just as good as phase shift. The wavelength change during measurements with wavelength shifting does not cause significant measurement errors. The wavefront error from Zygo measurement is a little different, which might be largely because the measured wavelength of Zygo IR interferometer is 1.5445 μm that offs a little from the claimed 1.55 μm. But the CGH is designed for 1.55 μm.

5.3.2 Element #2: Wedged Prism

Element #2 is a wedged prism with two concave surfaces . It is the easiest test among the three grism elements, because it does not include any diffractive surfaces. However, a CGH is still needed for correcting the huge aberration by tilted spherical surfaces. The setup for this CGH design has some similarities as Element #1. The CGH is in collimated space and the focus is used as a carrier. The main difference is that Element #2 test setup is designed in visible at λ = 632.8nm. Because no diffractive surface involved, use a HeNe interferometer is more convenient and provides more

precise wavefront measurement. The mounted Element #2 is show in Figure 10 in section 5.3.1. Figure15 shows the Element #2 test setup design and Figure 16 is the photo of lab setup.

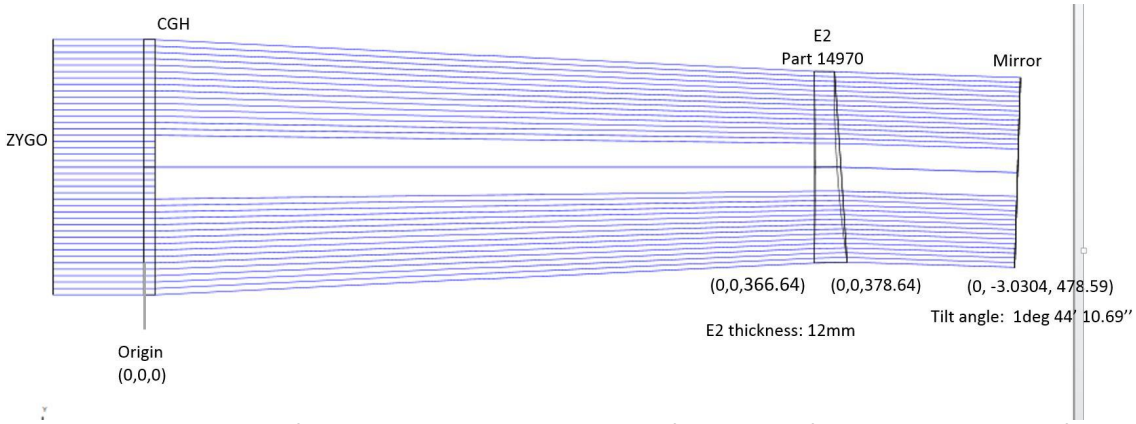


Figure 15. Element #2 wavefront test setup design. A collimated beam from Zygo interferometer has a wavelength of 632.8nm. The CGH sits in the collimated space. The focus carrier converges the beam after the CGH to compensate the negative power of E2, and provide wavefront correction at the same time. The light coming out of the E2 is collimated again but slightly tilted. The return mirror is aligned nominal to the beam for returning the beam back to interferometer.

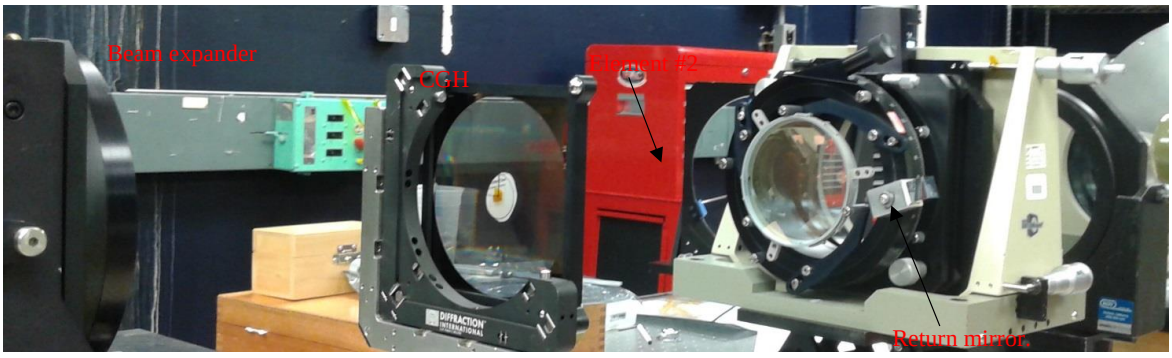


Figure 16. Real test setup of Figure 15. The theodolite behind the return mirror was used help align the various parts of the setup with respect to each other.

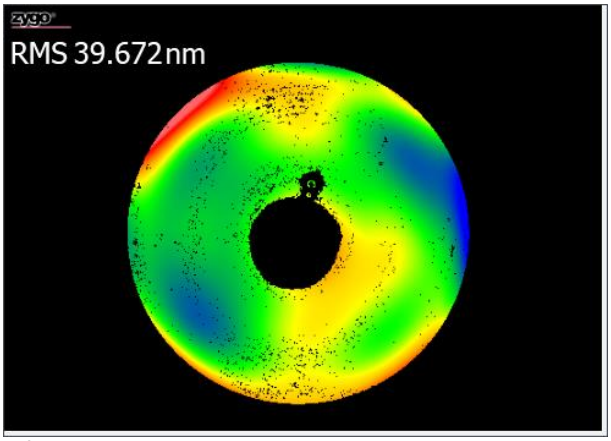


Figure 17. Measured wavefront error is < 40nm RMS. It meets the requirement.

Figure 17 is the measured wavefront error from the test setup in Figure xxx. The total wavefront error is 39.2nm, which meets our error budget. For this particular element, we also performed cryo test at 170ºK to verify the stress induced by the mounting ring. The test result is positive. The details will be published in a different paper.

5.3.3 Element #3: Grating

The grating is the last element to be tested. Its picture is shown in Figure 10 in section 5.3.1. The test setup is very similar to Element #1. Again, the CGH is placed behind a transmission flat into the collimated test beam. The CGH is designed to ensure that the beam is collimated and aberration free after Element #3, the wavelength of the interferometer is 1.55 μm. Figure 18 shows the measured wavefront map. Again, it was measured in three different ways: using the NIST IR^3 interferometer with mechanical phase shifting and wavelength shifting, and a Zygo IR interferometer with wavelength shifting only. The measurement result is shown in Table 4.

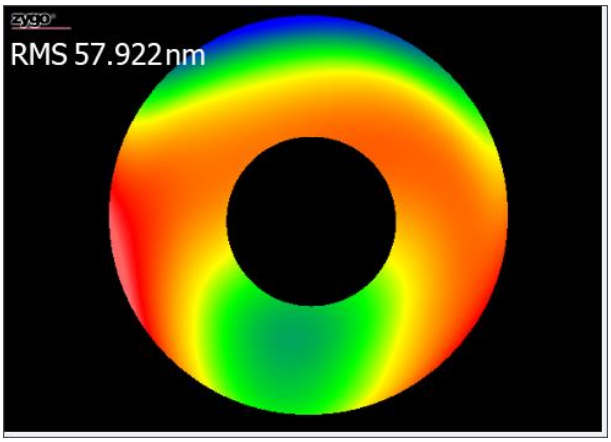


Figure 18. Wavefront map of Element #3 that is measured by NIST phase shift interferometer. The raw wavefront data is then processed using Zygo software MX.

Table 4. Measured Element #3 wavefront error with standard uncertainty using 3 different methods.

	NIST IR^3 piezo-shifting	NIST IR^3 wavelength shifting	Zygo wavelength shifting
Wavefront error (RMS)	58 nm	58 nm	55 nm

Finally, we would like to note that all three grism elements have acceptable wavefront error. We also like to note that all CGHs used for grism element tests are designed and built in-house. The NanoFab at NISTs Center for Nanoscale Science and Technology was used to plasma etch the amplitude CGH into phase CGH. This has not only greatly shortened test turnaround time, but also provides much more flexibility to accommodate our existing facility and equipment.

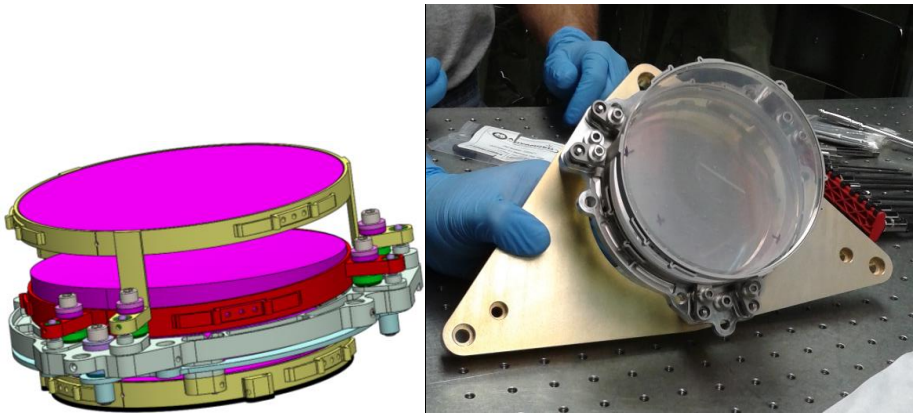
6. GRISM MECHANICAL DESIGN AND ASSEMBLY

The mounts manufactured to hold the grism elements were made of invar. This material was selected to minimize the wavefront distortion at 170°K, the operational temperature of the instrument.

Each individual element was mounted onto a ring structure which had three flexure points which made contact with the optic. To mount the elements epoxy Hysol EA 9309.2 was placed on each flexure and was left to cure for 5 days while the optic rested on a Teflon mount. Afterwards, these three ring mounts were mounted onto a cell which had nominal pin positions that allowed to position the elements in a nominal location. However, to mount the cell onto a tip/tilt mount for measurement purposes, a triangular integration deck was used, which is shown in Figure 19(b).

The individual ring structure mechanical design was previously tested in a similar environment for another project which also had to control the CTE mismatch which can introduce wavefront degradation.

Five degrees of freedom were designed into two of the three elements to provide some adjustment once mounted and assembled. Once assembled, the need to for adjusting the third element occurred, and parts were modified to accommodate from this.
To test the performance of the mount on the elements, Element #2 after mounting was cycled twice from 270°K to 160°K. During the cryogenic cycles one of the surfaces of Element #2 was measured in reflection through a window. After the cycles were completed, the measured wavefront was mostly undistorted, notifying a change of approximately 15nm in wavefront rms error, proving that the mount and epoxy cause low distortion. The CAD model of the grism prototype with its flexures is shown in Figure 19(a).



(a) (b)

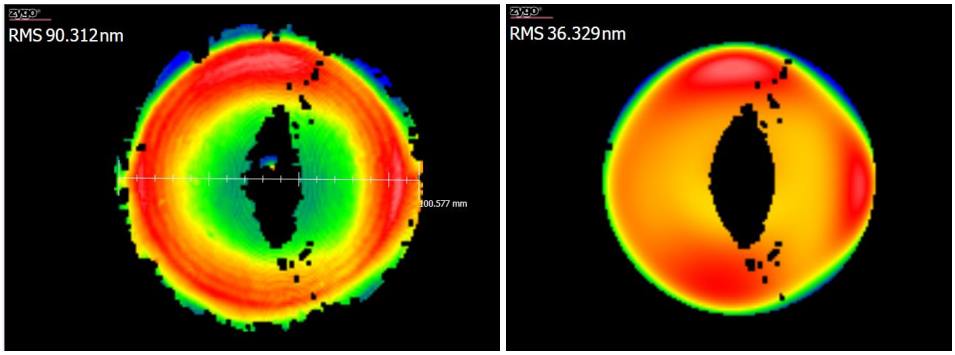
Figure 19. (a) CAD model of the GRISM assembly in its invar mounts. (b) Assembled grism prototype on integration deck.

7. GRISM ASEMBLY TEST AND PRELIMINARY RESULTS

After the elements were assembled, as Figure 19(b) shows, they were mounted onto the tip/tilt mount, the assembly was then placed on a hexapod in front of an infrared interferometer in non-collimated space, similar to the actual configuration of the grism in the instrument.
The testing configuration consisted of a wavelength shifting infrared interferometer, with a 1544.5nm wavelength (measured with a wavelength meter), an f/7.2 transmission sphere, the grism assembly and a spherical mirror that would allow to double pass through the grism (shown from left to right in Figure 20). The setup initially started by aligning the sphere to the interferometer and then introducing the grism in the path, since the assembly does not have power. As mentioned in the previous section the mount had some degrees of freedom that would allow for clocking (~0.4deg) and decenter adjustment (~650microns), which proved to be helpful since the alignment of the elements is sensitive to the clocking orientation one element has with respect to the other. After doing a sensitivity analysis, the elements were positioned at optimized locations which improved the wavefront measurement. The entire assembly was placed on a hexapod that allowed easy and accurate movement of the grism when varying the different field positions.

Figure 20. Test layout of the grism assembly in front of an IR interferometer with a flag indicating the focus position of the transmission sphere.

The preliminary unprocessed result of the aligned grism assembly is ~90nm WF rms at the on-axis field position. The wavefront is shown in Figure 21(a). The dark portion in the center of this figure shows significant drop out due to the grating of Element #1 that prints through the interferogram. However the processed measurement result (shown on Figure 21(b)) still validates the performance of the grism because the actual instrument has a pupil mask in front of the assembly that blocks out the center portion of it anyway. To obtain the processed wavefront, the high frequency errors were removed, the power term was removed also and inner and outer masks were used.
It is important to note that these results are preliminary and are only for the on-axis field position, the off-axis positions still have to be measured.
According to preliminary Zemax models which included the measured performance of each of the independent elements, the calculated wavefront performance should be <100nm rms. The allotted budget for the prototype assembly is ~115nm rms. Additional measurements, modeling and analysis are still required, however, these results are preliminary but encouraging.



(a) (b)

Figure 21. (a) Raw measurement of the grism assembly. (b) Processed wavefront of the grism assembly.

8. CONCLUSION

We have successfully designed, fabricated, assembled, and tested the cycle 5 prototype grism assembly. To date, we have verified the wavefront of each element and whole grism assembly. All elements meet specification. We are really happy to report that the performance of whole assembly at the only field position we have measured is exceed

Content, David; Dominguez, Margaret; Emmett, Thomas; Gong, Qian; Griesmann, Ulf; Kruk, Jeffrey; Marx, Catherine; Pasquale, Bert; Wallace, Thomas.
"Wide-Field InfraRed Survey Telescope (WFIRST) Slitless Spectrometer: Design, Prototype, and Results."
Paper presented at SPIE Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 26, 2016 - July 1, 2016.

specification. Besides the wavefront error, the diffraction efficiency from both Element #1 and Element #3 has almost reached theoretical prediction. This implies that the ghost from unwanted order is minimal. Our high fringe contrast during the interferometer measurement indicates the same result. Our plan next is to build a single field WFIRST simulator to perform the spectral test: to verify spectral resolution, grism spectral response (related to diffraction efficiency versus wavelength), as well as ghost from unwanted orders. The simulator moves to different positions and tilted to corresponding angles for mapping the FOV. Based on the experience and lesson learned from the prototype development, we'll start the phase A slitless spectrometer design study soon.

Disclaimer: The full description of the procedures used in this paper requires the identification of certain commercial products and their suppliers. The inclusion of such information should in no way be construed as indicating that such products or suppliers are endorsed by NASA or NIST, or are recommended by NASA or NIST, or that they are necessarily the best materials or suppliers for the purposes described.

ACKNOWLEDGMENTS

Our research was in part performed using resources provided by the NanoFab of the NIST Center of Nanoscale Science and Technology.

REFERENCES

[1] Scoville, N. etc. "The Cosmic Evolution Survey (COSMOS) - Overview", December, 2006 <http://arxiv.org/pdf/astro-ph/0612305.pdf>
[2] Gehrels, N., Spergel, D., Melton, M., and Grady, K., "WFIRST-AFTA Science Definition Team Final Report", (13 February, 2015)
[3] Gong, Q., Content, D., Kruk, J., Pasquale B., and Wallace, T., "DESIGN, FABRICATION, AND TEST OF WFIRST/AFTA GRISM ASSEMBLY", NASA New Technology Report (NTR#: GSC-17334-1)
[4] "Wide Field Camera 3 Instrument Handbook", Space Telescope Science Institute, January 2016 < http://www.stsci.edu/hst/wfc3/documents/handbooks/currentIHB/wfc3_ihb.pdf>
[5] "Near Infrared Camera and Multi-Object Spectrometer Instrument Handbook", Space Telescope Science Institute, http://www.stsci.edu/hst/nicmos/documents/handbooks/current_NEW/toc.html
[6] Swanson, G. J., "Binary Optics Technology: Theoretical Limits on the Diffraction Efficiency of Multilevel Diffraction Optical Elements", MIT Lincoln Lab Technical report, Pg. 11, (1 march, 1991)
[7] Malacara, D. [Optical Shop Testing], Wiley-Interscience, A John Wiley & Sons, Inc., Publication, third edition, 474-488 (2007)
[8] Malacara, D. [Optical Shop Testing], Wiley-Interscience, A John Wiley & Sons, Inc., Publication, third edition, 547-666 (2007)
[9] U. Griesmann, Q. Wang, M. Tricard, M. Dumas, and C. Hall, "Manufacture and Metrology of 300 mm silicon wafers with Ultralow Thickness Variation", AIP Conf. Proc. 93, 105-110, (2007)

The UT1 and UTC Time Services Provided By The National Institute of Standards and Technology

Judah Levine
Time and Frequency Division
NIST
Boulder, Colorado 80305
Judah.Levine@nist.gov

Abstract

I will describe the network-based time services provided by the Time and Frequency Division of NIST. The network service is realized with approximately 20 servers located at many locations in the US. The servers receive approximately 160 000 requests per second for time information in a number of different formats. The Network Time Protocol (NTP) receives approximately 98% of these requests. In addition to the servers that transmit UTC time, we also operate a single server that transmits UT1 time in NTP format. The UT1 server implements the offset between UT1 and UTC based the data in IERS Schedule A. The accuracy of the extrapolation is better than 4 ms; the accuracy of the time received by a user will depend on the stability of the network connection and is generally better than 8 ms. I will also describe the plan to improve the accuracy of the time service. This upgrade should be completed in the next few months.

Introduction

I will describe the clock ensemble that is used to generate the NIST time scales UTC(NIST) and TA(NIST). The atomic time scale, TA(NIST), was synchronized to TAI in 1978 when it was first implemented as a real-time scale. It has been free-running since that time. The UTC(NIST) time scale is realized as an offset from TA(NIST). It is steered towards UTC as computed by the International Bureau of Weights and Measures (BIPM). The steering is accomplished by automated frequency adjustments to the output phase stepper as I will describe below in more detail. The clocks are not steered.

The digital time services provided by NIST respond to requests in several different formats that are received over dial-up telephone lines and the public Internet. The servers currently receive approximately 220 000 requests per second or about 2.2×10^{10} requests per day. About 98% of these requests are in the Network Time Protocol (NTP) format [1,2], and the remainder are in several other older formats, which are supported for legacy applications. The number of requests continues to increase at a rate of about 4% per month, and most of this growth is in NTP-format requests.

The NIST Time scales

The NIST time scale [3] uses an ensemble of commercial high-performance cesium standards and hydrogen masers. The input to the ensemble calculation is an array of time-difference measurements between one of the clocks in the ensemble, which is designated as the reference clock for the measurement process, and all of the other clocks. The time differences are measured every 12 minutes. The reference clock for the measurement process is chosen for its reliability and stability, and has no special role in the ensemble calculation. The clock selected to be the reference clock also has no special role in generating the UTC(NIST) output signal.

Levine, Judah.
"The UT1 and UTC Time Services Provided by the National Institute of Standards and Technology."
Paper presented at The Science of Time: Time in Astronomy & Society, Past, Present and Future, Cambridge, MA, United States. June 5, 2016 - June 9, 2016.

The time-difference measurements are realized with a dual-mixer system [4]. The output from each clock at 5 MHz is mixed with a frequency of 5 MHz – 10 Hz, which is synthesized from the reference clock of the measurement process. The time difference of each clock is measured as the phase differences between the 10 Hz beat frequency generated by the mixing process and the 10 Hz frequency generated by the reference clock channel. The down-conversion of the 5 MHz input signals to 10 Hz increases the resolution of the measurement process by a factor of 5×10^5, so that a measurement resolution of 10^{-7} of a cycle at 10 Hz translates into an effective resolution of less than 1 ps at the input frequency of 5 MHz. The stability and accuracy of the time-difference data is limited by the corresponding parameters of the front-end mixer, and is typically a few ps. This performance is routinely verified by measuring the time difference between a single clock connected to two measurement channels. The differential nature of the time-difference measurements attenuates any common-mode variation in the characteristics of the measurement channels.

The atomic time scale TA(NIST) is computed as the weighted average of these time differences. The weight assigned to each clock is derived from its stability, which is calculated as the RMS residuals of the time difference between the time of the clock after its deterministic parameters have been removed and the time of the ensemble averaged over the previous 30 days. This method over-estimates the stability of a clock, since the clock is also a member of the ensemble that is used to evaluate its stability. This correlation is attenuated by decreasing the weight of a high-weight clock [5] and also by administratively limiting the weight of any clock to 30% of the ensemble. (Most time scale algorithms have similar limits.) Since the weights of all of the clocks must sum to 100%, limiting the weight of a good clock implicitly increases the weights of clocks that have poorer stability. Therefore, the administrative limitation on the maximum weight of any clock implicitly degrades the stability of the ensemble.

The frequency of TA(NIST) was set equal to the frequency of the International Atomic Time (TAI) frequency when the dual-mixer system was inaugurated, in 1978, and it has slowly drifted upwards in frequency since that time. The current frequency of TA(NIST) is greater than the frequency of TAI by approximately 37 ns/day (a fractional frequency of approximately 4.5×10^{-13}).

The UTC(NIST) time scale is derived from TA(NIST) by an offset in time and frequency that is calculated administratively. The offset is a piece-wise linear function that has no time steps. That is, changes to the offset equation are realized as changes in frequency only. The origin time of each of the linear steering equations is the ending time of the previous function. In other words, the origin time of each steering equation is the integral of all of the previous frequency offsets from 1978 when the system was initialized to the current time. The steering frequency has been negative for many years to remove the positive frequency offset of TA(NIST) that was described in the previous paragraph. The current and proposed future steering equations are published in the periodic Bulletin of the NIST Time and Frequency Division. [6]

The steering adjustments are computed from the data in the BIPM Circular T [7], which gives the difference between the UTC time scale computed by the BIPM and the time scales of all of the contributing laboratories. The Circular T for any month is typically published about the 10^{th} day of the following month. The BIPM also computes a more rapid version of UTC [8]. The rapid version, UTCr, provides daily estimates of the same data as in Circular T with a delay of one week. The results of the UTCr calculation are published on Wednesday afternoon (about 1700 UTC) for the period ending on the previous Monday. Starting in January, 2016, the data from UTCr were also used to compute a steering adjustment for UTC(NIST). This was done to improve the short-term accuracy of the UTC(NIST) time scale. The BIPM data for UTC-UTC(NIST) and UTCr-UTC(NIST) are shown in fig. 1.

After each 12-minute time scale computation, the software computes the offset in time and in frequency of UTC(NIST) with respect to TA(NIST) based on the appropriate steering equation, and applies these parameters to a hardware phase-stepper that produces the requested output signals at 5

MHz and 1 Hz. The reference signal for the phase stepper is derived from one of the clocks in the ensemble. The signal from the reference clock itself is not steered. The reference clock for the phase stepper has no special significance in the time scale algorithm, and is chosen mostly for its reliability and longevity. In general, the reference signal for the phase stepper is derived from one of the hydrogen masers in the ensemble, because these signals have the best short term frequency stability. The 5 MHz output signal from the phase stepper is fed back into the time scale measurement system, where it is measured along with all of the other clocks in the ensemble. However, the signal from the phase stepper is not included in the ensemble calculation, and its time difference with respect to the reference clock of the measurement system is used to verify that the requested steering offsets have been correctly applied. In normal operation, the difference between the physical output of the phase stepper and the equation that defines UTC(NIST) for that epoch is a few ps, and is well characterized as white phase noise.

The output signals from the phase stepper are used as the reference signal for all of the NIST time services. They are also used to control the link between NIST and the BIPM through PTB (The Physikalisch Technische Bundesanstalt, which is the National Metrology Institute of Germany) . This link is used by the BIPM to compute the values for UTC-UTC(NIST) that are reported in the monthly Circular T and in the more rapid UTCr data. The link between NIST and PTB is implemented using the two-way message exchange protocol, and this method is also used to support the NIST digital time services. I will describe this method in the next section.

The two-way method

Estimating the delay in transmitting messages from the source to the user is usually the most important and most difficult aspect of any time distribution system. The two-way method, in which the one-way delay is estimated as one-half of a measurement of the round-trip value, is widely used for this purpose.

The accuracy of the two-way method depends on the assumption that the delay is symmetric – that the inbound and outbound delays are equal. The inbound and outbound delays are comprised of two contributions: the delays through the station hardware at each end-point and the delay through the channel itself. The delays through the station hardware can be measured in principle, and any asymmetries can be calibrated and removed administratively. The uncorrelated variations in the inbound and outbound delays in the station hardware must be minimized as much as possible. A residual admittance to fluctuations in the ambient temperature is a common problem, and variations in the ambient temperature must be minimized for that reason.

The symmetry of the channel delay is a more difficult problem, because it is usually not under the control of the operators at either station. The symmetry requirement is not too difficult to realize with simple physical circuits, but it is much more difficult to do so when the circuit is realized as a packet-based communications channel with intervening routers, switches, and gateways. The stability of the delay in the satellite transponder, which is used in the link between NIST and PTB described in the previous section, presents a similar problem. The queueing delays in these elements often depend on the traffic load, especially when that load is asymmetric, as is often the case with systems that provide web-based services or streaming audio or video information. Even under the best of circumstances, an asymmetry of a few percent of the total round-trip delay is quite common in packet-switched networks, and this asymmetry often sets the limit on the accuracy of any distribution method that uses packet-switched networks such as the public Internet. Since the round-trip delay is often on the order of 100 ms in a packet-switched network, these networks typically provide accuracy that is often not much better than a few milliseconds. This limit has nothing to do with the format of the messages or the protocol

that is used to transfer the information, and there is no advantage of one protocol over another to the extent that the accuracy is limited by this consideration.

It is possible to mitigate the impact of the asymmetric delay contributions of routers, switches, and gateways in the communications channel by special purpose hardware at each one of these network elements. [9] The special-purpose hardware bridges the network element and measures the latency delay of a message passing through the device. The details of how the latency is measured vary somewhat from one implementation to another, but the important point is that this technique removes one of the primary limitations to the accuracy of packet-based time distribution. Unfortunately, these extra devices are not present in the general public Internet, so that the improvement in accuracy is not usually possible when the public Internet is used for time distribution.

The NIST time servers

The Time and Frequency Division of NIST operates a number of network-based time servers that are synchronized to the NIST clock ensemble. The servers that are located at the NIST facilities in Colorado are synchronized by a direct hard-wired connection between the 1 pps signal that realizes UTC(NIST) and the server hardware. The accuracy of the time at the server is a few micro-seconds. The servers that are located at other locations (see the Internet Time Service list at http:// tf.nist.gov for a list of the locations) are synchronized by a telephone connection between the server and the NIST clock ensemble in Boulder. The connection uses the ACTS protocol [10], which transmits time over dial-up telephone lines. The delay in the telephone connection is measured by using the two-way protocol described above, and the accuracy of the time at the server is typically a few milliseconds.

The servers respond to time requests received over the public Internet in a number of different formats. The most common request is in the Network Time Protocol (NTP) format. This format implements the two-way method for estimating the transmission delay, and the accuracy of the messages is limited by the considerations that I discussed in the previous section.

There are three types of NTP servers. Most of the NIST servers are synchronized to UTC(NIST)and respond to requests by using the standard NTP message format. The second type of server also transmits time data based on UTC(NIST) but includes support for message authentication based on the MD5 hash code. There are currently approximately 400 users of this service, and each one has a unique key. The key is combined with the standard NTP request and reply packets and the hash function of the combination is computed by the sender and appended to the message. The receiver computes the hash function of the message and compares the computed result with the transmitted version. The message is accepted only if the computed hash function agrees with the one in the message. Since the key parameter is a secret known only to NIST and the single registered user, the authentication process ensures that the inbound and outbound messages were not sent by a rogue third party and were not modified in transmit. The authentication process does not affect the accuracy of the time transmitted in this way, since the accuracy is limited by symmetry of the path and is not affected by the contents of the message.

The third type of NTP server transmits UT1 time in the standard NTP format. The UT1 time is computed by adding the difference between UT1 and UTC published by the International Earth Rotation and Reference Service (IERS) in Schedule A [11]. The schedule is published every week, and contains estimates of the UT1-UTC time difference for about a year in the future. The accuracy of the prediction for 30 days in the future is approximately 4 ms, and this is comparable to the accuracy of the NTP time transfer process as I have described above. Therefore, the values used by the time server are updated every few weeks, and the longer term predictions in each schedule are never used.

There are currently 16 registered users of this service. The service is limited to registered users so that the general user, who may not understand the difference between UTC and UT1, will not connect to the server by mistake and receive a time message that can be "wrong" by up to 0.9 s.

In addition to responding to requests in the NTP format, the servers also respond to requests in the DAYTIME [12] and TIME protocols [13]. These are simpler protocols that do not support the full two-way message exchange, and they are used when the client hardware is relatively simple and when the highest accuracy is not required. The NIST implementation of the DAYTIME protocol includes advance notice of the transitions to and from daylight saving time based on the US model, and also provides advance notice of leap seconds. The daylight saving time information is particularly useful for older systems that may not have any other source for this information or that may have the older, incorrect US transition dates. The TIME protocol is the simplest of all of the message formats, and is supported for legacy applications only. We do not recommend it for new applications because it has poor error-checking capabilities, supports no method for estimating the network delay, and provides none of the ancillary information of the DAYTIME format.

The NIST time servers receive approximately 220 000 requests per second for time in the various formats, and almost all of these requests (> 98%) are in the NTP format. The number of requests has been increasing at a rate of approximately 4% per month, and this rate of increase shows no sign of moderating. The increase in the number of requests since 2004 is shown in fig. 2. The figure shows that the number of requests in the DAYTIME and TIME formats has been decreasing since 2012, but there are still approximately 10^8 requests per day in these formats, or about 1 000 requests per second.

Traceability

The concept of traceability is important for all time services. A time measurement is traceable if there is an unbroken chain of measurements from the end user back to a national or international standard of time. [14] Each link in the measurement chain must be characterized by a delay and an uncertainty. In the US, traceability can be satisfied by a chain that leads back to UTC(NIST) or UTC(USNO).

In addition to the technical traceability requirement of the previous paragraph, some users are required to have legal traceability. In addition to the technical requirements for traceability, legal traceability adds requirements that can confirm technical traceability in a legal adversary proceeding. The details of how legal traceability is satisfied depend on the end-user, but generally include properly maintained and certified log files and other documentation that can confirm the proper operation of the synchronization process. A log file that contains only error information may be inadequate, because it can be completely empty when the system is working properly, and it can be hard to distinguish this case from a system that is totally inoperative.

Although a time signal may be generated by a traceable source such as a GPS satellite or a NIST-operated time server, the traceability of these signals generally does not extend to un-calibrated and unmonitored user equipment. That is, the traceability of a signal from a GPS satellite ends with the signal in space by default, and the default traceability of the time from a NIST time server ends at the server portal. The messages transmitted by third-party servers is generally not traceable.

The Leap Second problem

The current definitions of the length of the second in both the UTC and TAI time scales are based on the same value derived from the frequency of a hyperfine transition in the ground state of cesium 133 [15]. The length of the second that results from this definition is somewhat shorter than the length of the second of the UT1 time scale, which is derived from the rate of rotation of the Earth. The

difference between the two time scales has a systematic contribution whose magnitude is currently somewhat less than 1 s per year, and an additional contribution that is irregular and whose magnitude cannot be accurately characterized either deterministically or stochastically. The stochastic contribution to the length of the UT1 second could reverse the sign of the UTC-UT1 difference in principle, but this has not happened to date and is unlikely to happen in the future. The UTC time scale is used as a proxy for UT1 time in many applications ranging from every day timekeeping to the timing of events in astronomy, and maintaining the equivalence between UT1 and UTC is an important consideration in timekeeping.

In order to prevent the UTC and UT1 time scales from diverging, an extra "leap" second is added to UTC whenever the magnitude of the difference between UTC and UT1 approaches 0.9 s. [16] (As I mentioned in the previous paragraph, it could be necessary in principle to drop a second in UTC to preserve this equivalence, but this is unlikely in the foreseeable future.) This extra second is added after 23:59:59 UTC, usually on the last day of June or December. (The last days of other months are possible in principle but have never been used.). The name of this extra second is 23:59:60 UTC. The leap second is defined with respect to UTC, and it therefore occurs at different local times at different locations. For example, it occurs late in the afternoon in California and during the following morning in Asia and Australia. The second after the leap second is named 00:00:00 UTC of the next day.

Computer clocks and most digital systems keep time internally as the number of seconds that have elapsed since an origin time, which can differ from one system to another. The conversion between the count of the elapsed seconds and the time in the conventional year-month-day hour-minute-second representation is computed by the operating software whenever the time is requested. Additional offsets for the extra day associated with leap years, for conversion to the local time zone, and for daylight saving time are also applied by this application if necessary. The various offsets are either strict constants (as for the offset of the local time zone) or can be computed algorithmically (as for the offsets for a leap year or for the local version of daylight saving time). Both leap days and the daylight saving time transitions introduce discontinuities between a time interval computed from the count of the elapsed seconds maintained by the system internally and a simple computation of the same time interval using the hour-minute-second representation. This discontinuity can be particularly serious during the Fall transition away from daylight saving time, since the local time moves backwards by one hour during the transition and the same time stamps will be transmitted twice. There is often no indication whether the time reported by the system during this transition period is the old daylight time or the new standard time. The impact of these changes is mitigated by the fact that the change is made at 2 am Sunday local time.

The method of time-keeping in a digital system is therefore divided into two cooperating processes: a low-level application that receives and counts the periodic signals generated by the hardware, and a second higher-level application that "knows" how to convert this count to the conventional representation of hours-minutes-seconds. In general, the partition between the low-level and high-level processes is arbitrary because the application need not be concerned with the intermediate values transmitted between them. However, leap seconds are inserted on an irregular and unpredictable basis, and the times of both future and past leap seconds cannot be computed algorithmically. Therefore, there is no simple way of incorporating leap seconds into this deterministic pair of cooperating processes.

The solution used by the NIST time servers and by many other systems is to stop the hardware clock for one second at 23:59:59 UTC on the day of a leap second. The result is to use the time 23:59:59 UTC twice – once when that time occurs and a second time during the leap second. Neither the system applications nor the standard communications protocols have any provision for indicating that the second 23:59:59 time value is actually a leap second and should be assigned the time 23:59:60. This solution has a number of undesirable side-effects. The fact that there are two seconds with the same

name introduces confusion and can produce time stamps that appear to violate causality: 23:59:59.2 during the leap second actually occurred after 23:59:59.5 during the first second with this time stamp value, but the time stamps themselves indicate the opposite order. (The Fall transition from daylight saving time to standard time has the same problem with time stamps between 2 am and 3 am.)

This method of repeating 23:59:59 is probably the closest approximation to the definition of the leap second on systems that have no method of indicating the leap second event to applications and inserting a special leap-second indicator into communications protocols. The NIST DAYTIME protocol transmits time in the hour-minute-second format, and transmits the leap second time correctly as 23:59:60.

Although the method I have just described has undesirable side effects, there are other implementations that are worse. For example, a number of network time-service appliances add the leap second to the first second of the new day. That is, the time stamps in the vicinity of a leap second are 23:59:58, 23:59:59, 00:00:00, 00:00:00. The long term behavior of this method is correct, but it adds the leap second to the wrong day, and has a transient error of 1 second with respect to the definition. Other systems implement the leap second as a frequency adjustment over some time interval before (or after) the leap second. [17] This method introduces a varying time error on the order of 0.5 s with respect to the definition and a frequency error that depends on the interval over which the leap second is inserted. Since the interval for amortizing the leap second is not defined in any standard, different implementations of this method may produce time stamps in the vicinity of a leap second that disagree both with each other and with respect to UTC.

Finally, it is clear that all of these solutions introduce steps in both time interval and frequency in the vicinity of a leap second, and that physical processes do not stop during the leap second. The time scales of navigation systems do not use leap seconds for this reason, although each navigation system time typically includes the number of leap seconds that have been inserted from the start of the leap second system in 1972 to the instant that the navigation system time scale was initialized. Each navigation system time will therefore have a different offset from UTC for this reason, and all of these offsets will change whenever a new leap second is inserted. This multiplicity of time offsets from UTC is undesirable because it can be a source of confusion when time stamps are derived from the transmissions of navigation satellites without a clear indication of which time scale was used.

A number of proposals have been advanced for addressing the leap second problem, but a detailed discussion of them is outside of the scope of this document, which is focused on the NIST time services.

Improvements to the service

The Time and Frequency Division of NIST is addressing the continued increase in the number of requests for time services by upgrading the capacity of the servers and the network connections. We plan to have a number of very high capacity sites operating in the next few months. In addition, we are adding reference clocks to each of these sites to improve the accuracy of the time at the servers. We expect that the reference clocks at the remote sites will be within 10 – 20 ns of UTC(NIST), and we will consider offering services that can transmit the time to a user's system with an end-point accuracy of 100 ns or better.

Summary and conclusions

The NIST Time and Frequency Division operates an ensemble of time servers that respond to requests in a number of different standard formats. The servers receive approximately 220 000 requests per second, most of which are in the NTP format. The time of the servers is directly traceable to the

UTC(NIST) time scale, which is maintained by an ensemble of atomic clocks and hydrogen masers located at the NIST laboratory in Boulder, Colorado.

The time servers implement a leap second by transmitting the time corresponding to 23:59:59 a second time during the leap second event. This method is used because neither the system nor the NTP message format can represent the correct name of the leap second, which is 23:59:60. The DAYTIME format uses the hour-minute-second representation and transmits the correct representation of the leap second.

NIST operates a single server that transmits UT1 time in the standard NTP format. The internal time of the server is synchronized to UTC, and the difference between UT1 and UTC is added to the NTP response just before it is transmitted to the system that requested it. The UT1-UTC time difference is derived from IERS Schedule A. The difference published in this schedule for each day is used without interpolation, since the change in the time difference from one day to the next one is comparable to the accuracy of the NTP process, so that a typical user would probably not benefit from the increased accuracy provided by the interpolation process. There are currently 16 registered users of this service.

NIST also operates three time servers that authenticate the time stamps by computing the MD5 hash of the combination of a secret key and the standard time message. Each registered user of this service is given a unique key. There are currently approximately 400 users of this service. The authentication process ensures that the message originated from NIST and was not modified in transit. It does not affect the accuracy of the time service, which is limited by the symmetry of the network connection between the client and the server.

References

[1] D. L. Mills, Computer Network Time Synchronization, 2nd edition, CRC Press, Boca Raton, Florida, p64, ff.

[2] David L. Mills, RFC 1305, available from www.ietf.org/rfc/rfc1305.txt.

[3] J. Levine, The statistical modeling of atomic clocks and the design of time scales, Rev. Sci. Instr. Vol. 83, 012201-28, (2012).

[4] S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe and L. Erb, "Performance of an automated high accuracy phase measurement system,} Proc. 36th Annual Symposium on Frequency Control, Fort Monmouth, New Jerse, 1982. Reprinted in NIST Technical Note 1337, available from the publications list at tf.nist.gov.

[5] P. Tavella and C. Thomas, Comparative Study of Time Scale Algorithms, Metrologia, Vol. 28, pp. 57-63, 1991.

[6] The Steering data is available online at http://www.nist.gov/pml/div688/grp50/nistat1.cfm.

[7] Published monthly by the International Bureau of Weights and Measures and available from the BIPM web site at ftp://ftp2.bipm.org/pub/tai/publication/cirt/

[8] Details and data for all contributing laboratories are available on the BIPM web page at www.bipm.org.

[9] IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, IEEE Standard 1588, 2008, available from webstore.ansi.org.

[10] J. Levine, M. Weiss, D. D. Davis, D. W. Allan and D. B. Sullivan, The NIST Automated Computer Time Servcie, J. Res. NIST, vol. 94, pp 311-322, (1989).

[11] International Earth Rotation and Reference Service, http://datacenter.iers.org/eop/-/somos/5Rgv/latest/6.

[12] The NIST version of the DAYTIME protocol is described on Internet Time Service page at tf.nist.gov. See also, J. Postel, www.tfc-editor.org/info/rfc867, 1983.

[13] J. Postel, The Time Protocol, RFC 866, www.rfc-editor.org/info/rfc866, 1983.

[14] Barry N. Taylor and Chris E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note 1297, Washington, D.C., U. S. Government Printing Office, (1994).

[15] Le Systeme International d'Unites, 7th ed., Sevres, France, BIPM, 1998.

[16] Leap Seconds are described at www.nist.gov/pml/div68/leapseconds.cfm. See Also www.itu.int/en/ITU-R/Pages/default.aspx.

[17] https://googleblog.blogspot.com/2011/09/time-technology-and-leaping-seconds.html.

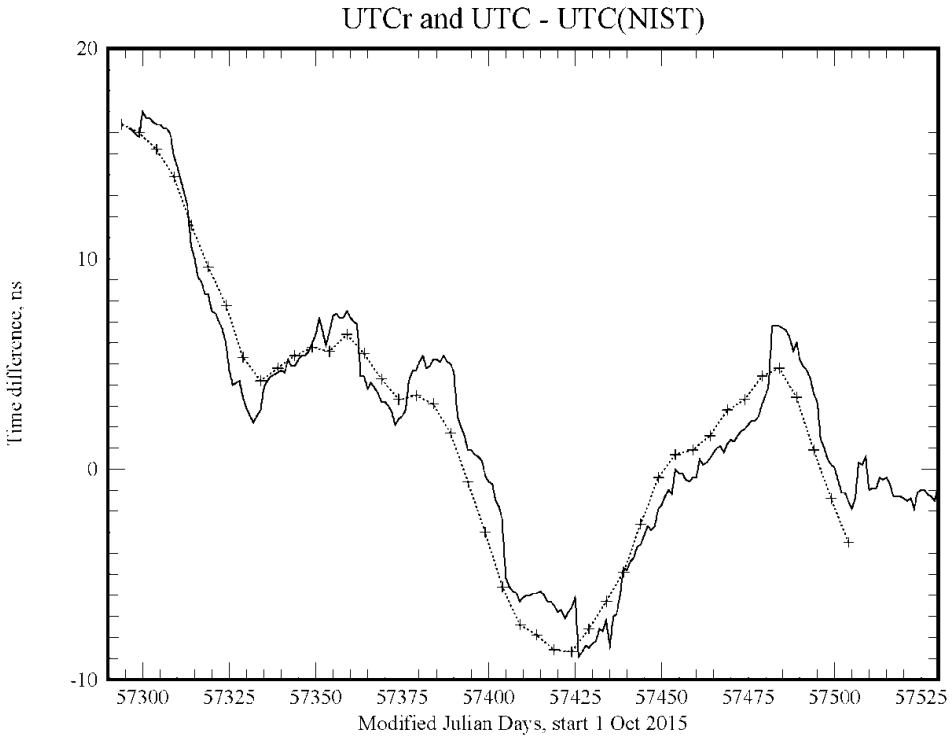


Fig. 1. The time differences in ns between the UTC and rapid UTCr time scales computed by the BIPM and UTC(NIST). The dotted line with the symbols shows the UTC time difference and the solid line is the UTCr time difference. The X axis is in Modified Julian Days, and the plot shows the time differences from October, 2015 to April, 2016.

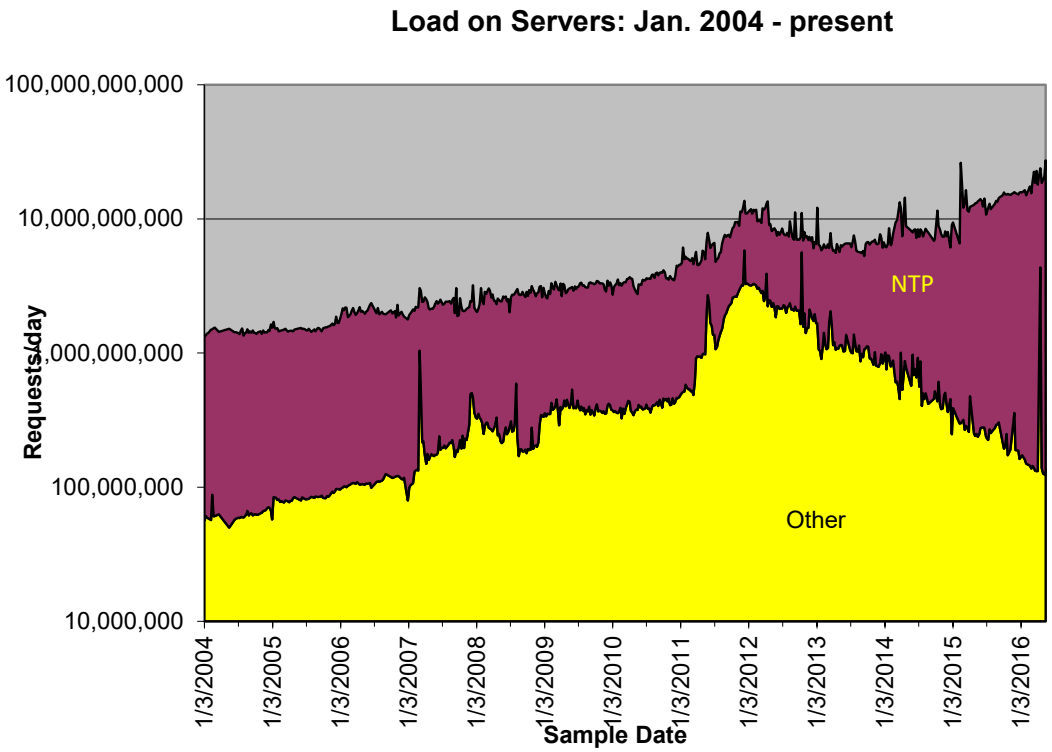


Fig. 2. The number of time requests received by all of the NIST time servers from January, 2004 to April, 2016. The logarithm of the number of requests as a function of time is displayed. The lower curve, labeled “other” is the number of requests in the TIME and DAYTIME formats, which the upper curve is the number of requests in NTP format.

Process improvements in the production of silicon immersion gratings

Cynthia B. Brooks[a], Benjamin Kidder[a], Michelle Grigas[a], Ulf Griesmann[b],
Daniel W. Wilson[c], Richard E. Muller[c], Daniel T. Jaffe[a]
[a]The University of Texas at Austin, Austin, TX, USA
[b]National Institute of Standards and Technology, Gaithersburg, MD, USA
[c]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

ABSTRACT

We have explored a number of lithographic techniques and improvements to produce the resist lines that then define the grating groove edges of silicon immersion gratings. In addition to our lithographic process using contact printing with photomasks, which is our primary technique for the production of immersion gratings, we explored two alternative fabrication methods, direct-write electron beam and photo-lithography. We have investigated the application of anti-reflection (AR) coatings during our contact printing lithography method to reduce the effect of Fizeau fringes produced by the contact of the photomask on the photoresist surface. This AR coating reduces the amplitude of the periodic errors by a factor of 1.5. Electron beam (e-beam) patterning allows us to manufacture gratings that can be used in first order, with groove spacing down to 0.5 micrometer or smaller (2,000 grooves/mm), but could require significant e-beam write times of up to one week to pattern a full-sized grating. The University of Texas at Austin silicon diffractive optics group is working with Jet Propulsion Laboratory to develop an alternate e-beam method that employs chromium liftoff to reduce the write time by a factor of 10. We are working with the National Institute of Standards and Technology using laser writing to explore the possibility of creating very high quality gratings without the errors introduced during the contact-printing step. Both e-beam and laser patterning bypass the contact photolithography step and directly write the lines in photoresist on our silicon substrates, but require increased cost, time, and process complexity.

Keywords: Immersion, diffraction, gratings, infrared, spectroscopy, lithography, silicon

1. INTRODUCTION

Silicon immersion gratings improve performance and reduce spectrograph size by taking advantage of the high index of refraction (3.4) of silicon to increase dispersion and allow extensive wavelength coverage at high resolution in a single exposure by permitting operation in high order. Silicon immersion gratings produced at the University of Texas at Austin provide the primary dispersive elements in the IGRINS instrument at the McDonald Observatory and the iShell instrument being built for the Infrared Telescope Facility on Mauna Kea, Hawaii, and will be used in the planned Giant Magellan Telescope Near-Infrared Spectrograph (GMTNIRS). To manufacture immersion gratings, we produce grooves in silicon using lithography followed by plasma etching and wet etching. Our current process uses contact photolithography to pattern ultraviolet (UV) sensitive photoresist as the initial processing step. We then transfer this pattern into a layer of silicon nitride that, in turn, serves as a hard mask during the potassium hydroxide etching of V-grooves into silicon. The initial patterning of the grooves in resist is the dominant step that determines the quality of the immersion grating. Contact lithography can introduce global uniformity errors from non-uniformity in the light source and periodic errors caused by Fizeau fringes in the contacting of the photomask to the substrate surface.

The manufacture of diffraction gratings requires the precise positioning of grooves over long distances. Immersion gratings, in particular, have very stringent manufacturing specifications because of the high index of refraction of silicon, which also gives immersion gratings their advantage. Marsh, *et al.* [1] described the groove placement requirements for silicon immersion gratings and Wang, *et al.* [2] interpreted these for an instrument with resolving power $R > 80{,}000$ at 1.6 μm. The combined large-scale wavefront error in the grating surface and entrance face must be smaller than $\lambda/4$ peak to valley to maintain resolving power and efficiency when operating close to the diffraction limit. To keep ghost levels less than 0.2%, periodic errors must have an amplitude less than 4 nm.

Brooks, Cynthia; Griesmann, Ulf; Grigas, Michelle; Jaffe, Daniel; Kidder, Benjamin; Muller, Richard; Wilson, Daniel.
"Process improvements in the production of silicon immersion gratings."
Paper presented at Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 25, 2016 - July 1, 2016.

As we push our immersion gratings to work at higher resolving power and shorter wavelengths, down to the silicon cutoff of 1.15 µm for use in instruments such as the planned GMTNIRS, these requirements become even stricter. To achieve diffraction-limited operation, the peak to valley groove placement error must be less than 85 nm over the entire grating and, to achieve ghost levels of not exceeding 0.1%, periodic errors must be maintained to below 2 nm in amplitude. In this developmental work, we endeavor to continually improve our grating lithography process by exploring alternatives to our standard contact printing process.

2. CONTACT PRINTING WITH ANTI-REFLECTIVE COATINGS

A problematic step in our contact printing process occurs during the actual contacting of the photomask to the photoresist on the polished silicon nitride/silicon substrate. As reported previously [3], the phase non-uniformity can be deconstructed into two components: large scale non-uniformity caused by imperfect UV illumination and smaller scale, residual non-uniformity due to Fizeau fringes that occur during imperfect alignment when contacting the photomask to the polished silicon substrate. We studied the addition of anti-reflective materials applied underneath our standard photoresist to reduce the contrast, and thus the magnitude, of these interference fringes.

We prepared two nitride over silicon wafers by coating the first a photoresist layer with 780 nm nominal thickness (Microposit S1800, manufactured by Shipley Company) alone and the second with 200 nm of an anti-reflective material (BARLi-II, manufactured by AZ Electronic Materials) underneath 780 nm of the same photoresist. We measured the reflectance of each using a Cary 5000 spectrophotometer. These measurements and the emission spectrum for the mercury lamp that serves as our UV source are plotted in Figure 1. The Microposit S1800 photoresist is designed to be exposed at 365 nm but is also sensitive at 405 nm [4]. Note that the reflectance is suppressed in the region used to expose photoresist. We predicted that the suppressed reflection would reduce the magnitude of the fringes produced during the contacting and exposure process.

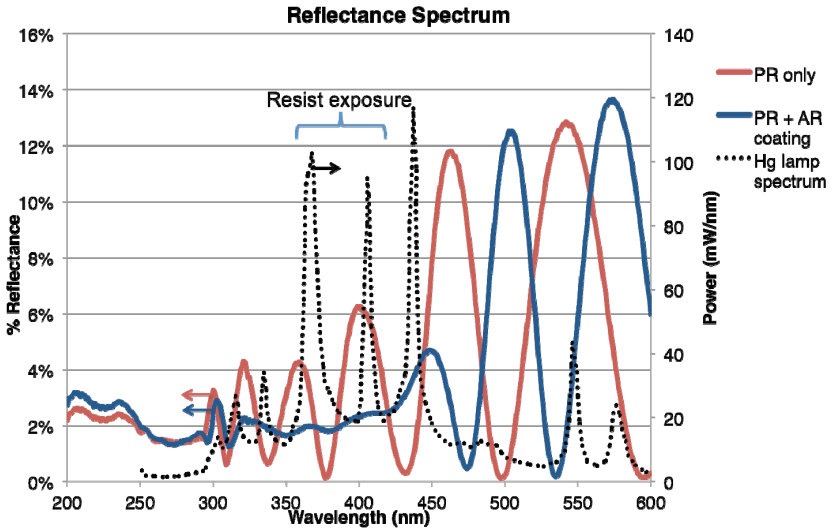


Figure 1. Reflectance spectrum at normal incidence measured on a Cary 5000 spectrophotometer for two conditions: photoresist (PR) only (red) and photoresist + anti-reflective coating (blue). For comparison, the dashed line is the mercury lamp spectrum that is used to expose the photoresist, which is sensitive to the 365 nm and 405 nm emission lines.

Because the BARLi-II coating is not developer soluble, it must be plasma-etched to expose the silicon nitride hard mask, as illustrated in the process flow in Figure 2. We developed the BARLi-II patterning process using thin silicon wafers that we could then cross-section and image with a scanning electron microscope, as shown in Figure 3. The photoresist and BARLi-II form the mask for plasma etching the silicon nitride.

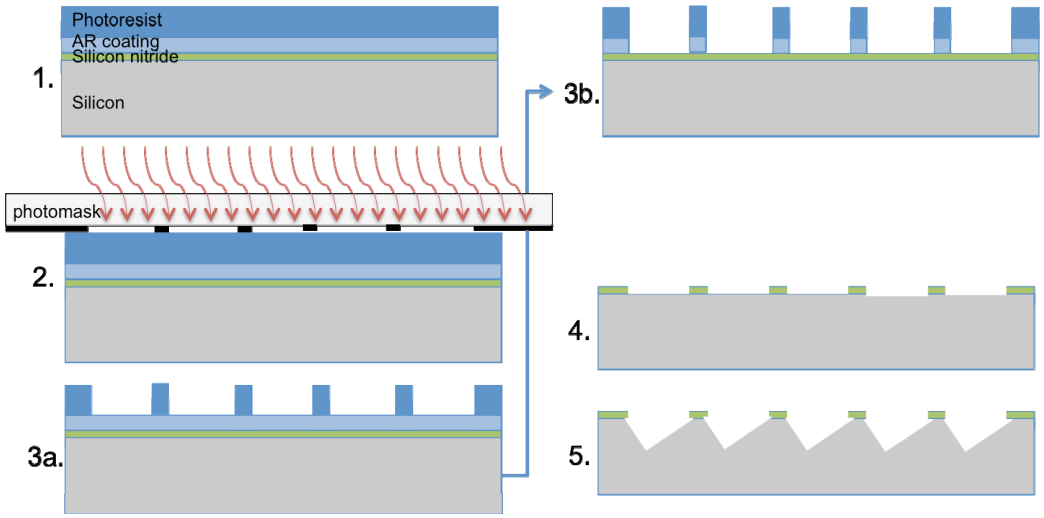


Figure 2. Process flow for manufacture of silicon immersion gratings using contact printing. (1) An anti-reflection coating is spun onto a silicon nitride over silicon substrate followed by a photoresist. (2) The photoresist is exposed through a photomask consisting of lines patterned in chromium. (3a) The photoresist is developed, with the exposed areas being washed away. (3b) The anti-reflective coating is plasma etched to expose the silicon nitride. (4) The silicon nitride mask is etched away in a fluorine-containing plasma and the photoresist and AR coating are removed. Finally, (5) V-grooves are wet etched in a potassium hydroxide solution.

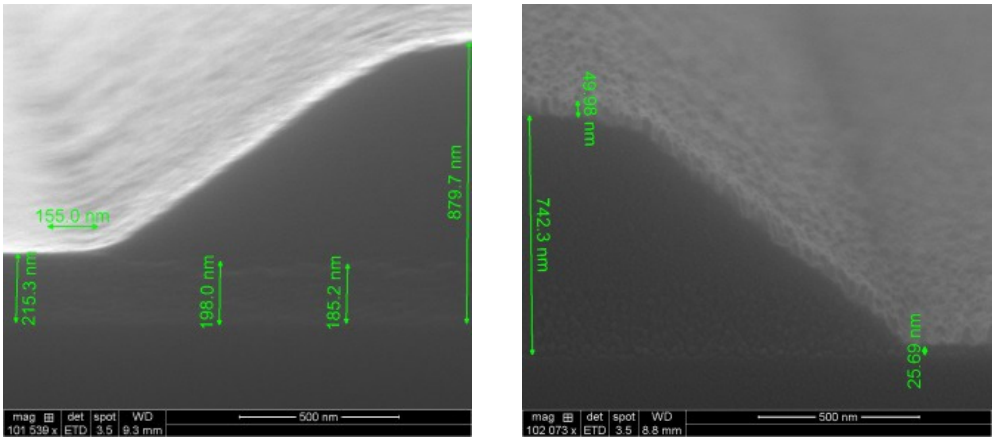
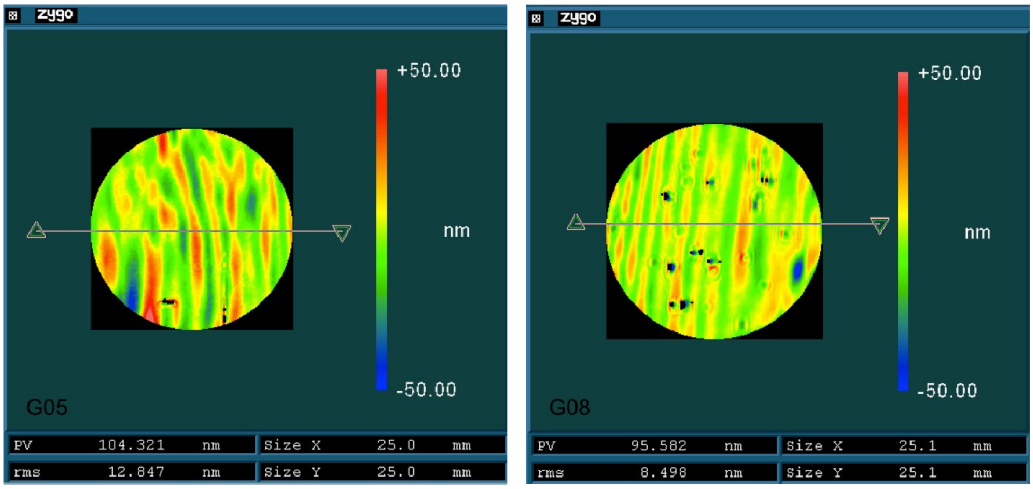


Figure 3. Cross-sectional scanning electron micrographs (SEM) of patterned wafers before (left) and after (right) BARLi-II etch. The BARLi-II thickness is ~200 nm and the thickness of the full resist over BARLi-II stack is ~880 nm. After etching, the stack is ~740 nm and ~25 nm of BARLi-II remains. This thin layer of BARLi-II will be etched away at the beginning of the subsequent silicon nitride etch process.

Using the described process, we deliberately patterned a grating with visible Fizeau fringes and produced a silicon immersion grating blazed at 71.6° (R3). We measured the phase error of light reflected by the front surface in Littrow configuration using a laser interferometer at 633 nm and modeled the first 36 Zernike components. These were subtracted from the full interferogram resulting in a residual error that is related to the fringe amplitude. This grating was compared to a previous grating manufactured without anti-reflective materials. Figure 4 shows that the root mean square (rms) of the fringe error was reduced from ~12.8 nm without anti-reflective layer to ~8.5 nm with the anti-reflective layer.



rms = 12.8 nm With photoresist rms = 8.5 nm With AR coating + photoresist

Figure 4. Residual error after subtracting the first 36 Zernike terms from the front surface interferogram of immersion gratings manufactured without (left) and with (right) anti-reflective coating. The rms value of the residual error was reduced from 12.8 nm to 8.5 nm.

3. ELECTRON BEAM PATTERNING

Electron beam lithography offers an alternative method of patterning lines. E-beam lithography is particularly well suited for patterning fine features. However, it can require prohibitively long write times for fine features over a large grating area. We studied three alternative methods of binary e-beam patterning. Those methods are compared in Table 1.

Table 1. Comparison of different methods of writing silicon gratings using electron beam patterning.

Method	Pros	Cons
Negative resist	Short e-beam write time because of smaller write area	New process requiring considerable process development Poor adhesion of resist
Positive resist with two pass patterning	Extensive experience with this kind of resist Write time improved over one pass patterning	Both passes must have same exposure Both passes must be perfectly aligned to each other
Positive resist with chromium liftoff	Short e-beam write time because of smaller write area Extensive experience with this kind of resist	Additional processing steps take extra time and cost More steps where something could go wrong

3.1 E-beam lithography with negative photoresist

As we have reported previously [5], the established process for writing lines in e-beam resists is to use a positive resist, illustrated in Figure 5, left side. With a positive resist, the electron beam is used to pattern the area that will be washed away during the develop step. With a negative resist (Figure 5, right side), the area that is *not* exposed to the electron beam is washed away during developing. In this case, only 10% of the area must be written resulting in a significant savings in e-beam write time and an expected improvement in reliability.

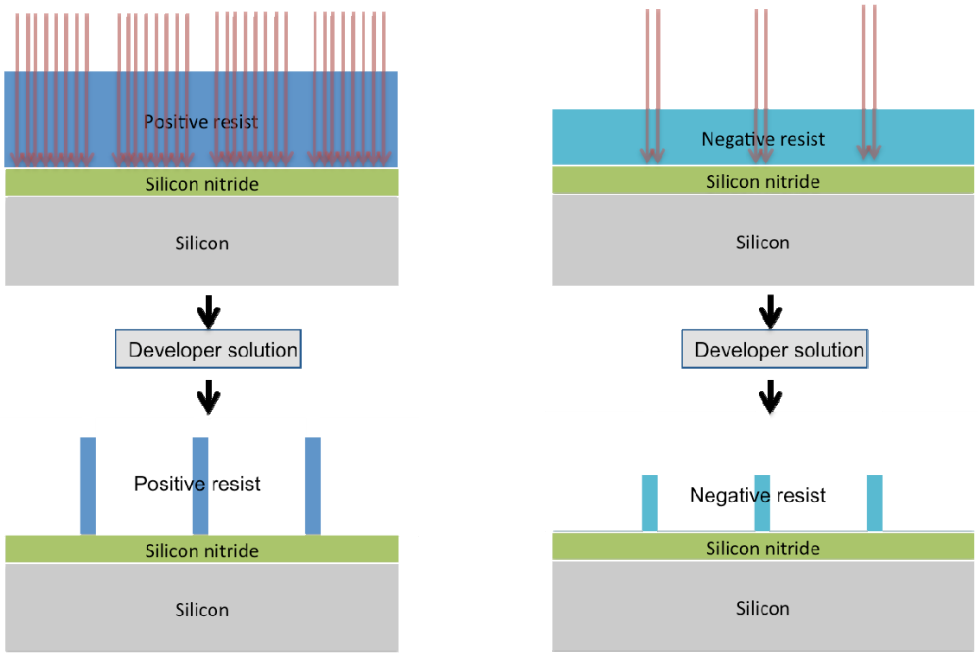


Figure 5. Comparison of positive e-beam resists (left) and negative e-beam resists (right). A positive resist is developed out in the region that is exposed to the electron beam. A negative resist is developed out in the region not exposed to the electron beam.

We chose the negative tone photoresist ma-N 2405 manufactured by Micro Resist Technology, wrote a very small test area using the negative resist and patterned it through plasma etch and KOH groove etch. Microscopically, the initial results were very promising and we were able to successfully pattern very small v-grooves in silicon, as shown in the SEM in Figure 6, left side. However, macroscopically we had poor adhesion of the resist to the substrate as illustrated in the optical microscope image on the right side of the same figure. This issue likely could have been overcome with iterative process development, but since we frequently pattern thick substrates that can be difficult to spin-coat and bake reliably, we decided to pursue processes based on positive-tone resists with better adhesion.

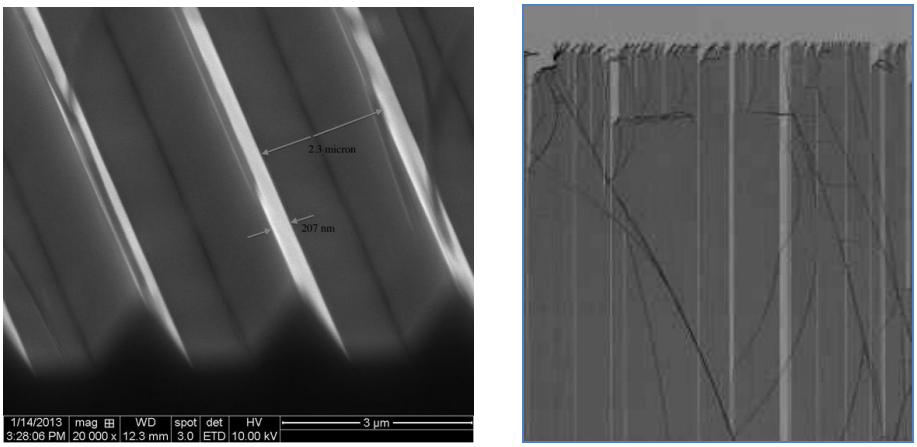


Figure 6. (left) Scanning electron micrograph of test wafer e-beam patterned using negative resist. (right) Optical microscope image showing loss of adhesion and lift off of lines over a larger area of the pattern.

3.2 E-beam lithography with positive resist using two-stage writing

To mitigate the long write times with positive resist, we developed a two-step patterning process with a high-resolution, low-current write to precisely define the line edges, followed by a low-resolution, high current step to write the bulk of the line. Figure 7 illustrates the planned write strategy for the two-pass process.

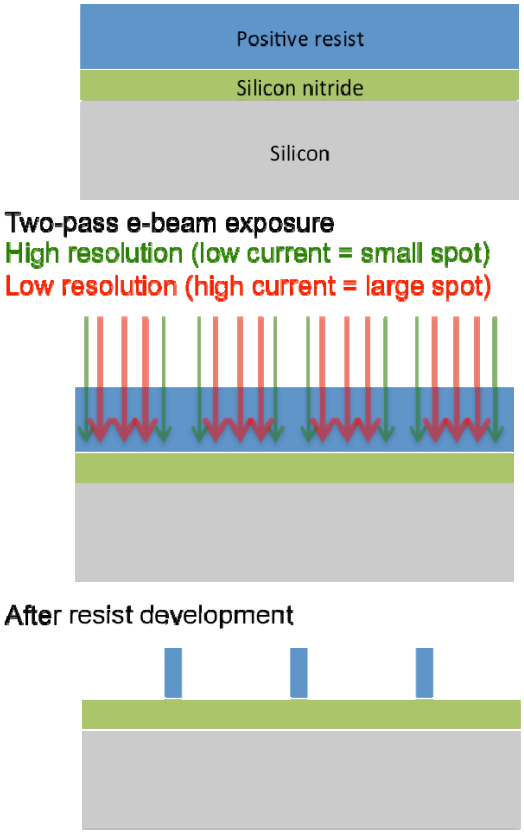


Figure 7. Process flow for two-step, positive resist patterning of immersion gratings. The high current, large spot size is used to write the bulk of the line. The low current, small spot size is used to precisely define the line edges.

The two-pass strategy requires very precise control of the e-beam dose and the physical alignment of the high and low current passes. Figure 8, left side, shows early results of the two-pass strategy. This image shows that the dose is correct for the high-resolution pass but because it is under dosed in the low-resolution pass the resist is not fully developed out. Figure 8, right side, is a Vernier overlay pattern designed to test the alignment of the high and low current passes. The best-aligned row, indicated by the red arrow, shows that there is still a 200 nm drift between the two passes. This is too much for our scheme to work because the overlap between high current and low current patterns is only about 40 nm. For this experiment, we only used the alignment marks on the e-beam stage to align the low-current and high-current exposures. Better results would have been achieved with alignment marks fabricated on the substrate, but that would have required an extra full process.

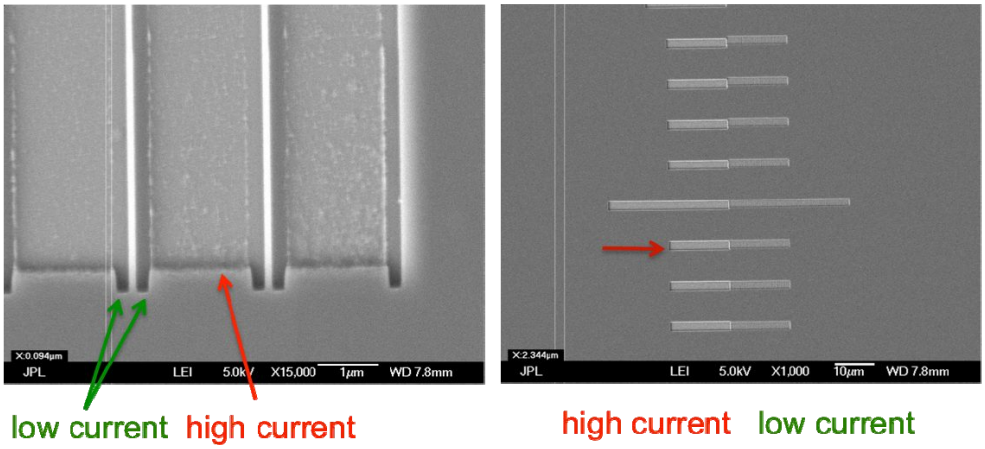


Figure 8. SEM images of results of the two-pass write strategy. (Left) Patterns were misaligned in the y-direction but are good in the x-direction. However, the high current pattern was underdosed and did not fully develop. (Right) Vernier pattern shows a 200 nm drift between high current and low current exposures. The resist was ZEP-520A manufactured by ZEON Chemicals.

3.3 E-beam lithography with positive resist using chromium liftoff technique

We devised a third strategy that would allow writing of the small lines with similar write time as the negative resist but would use the well-established positive resists ZEP 520A (ZEON Chemicals) and PMGI (Microchem Corp.). This method [6], commonly referred to as liftoff, relies on adding additional steps to the process as illustrated in Figure 9. In this process, spaces are written by e-beam and developed that are the opposite of the final lines needed in the part. Because our lines are narrow, this greatly reduces the write time allowing us to use the e-beam tool optimally. A thin layer of chromium is deposited over the entire patterned surface and the e-beam resist is dissolved away "lifting off" the chromium above it and leaving tiny lines. These lines form the mask used to etch the silicon nitride away. The chromium is then etched off and the silicon nitride forms the mask for etching the v-grooves of the grating per our usual process. Figure 10 shows SEMs of tests of the process using thin silicon wafers. The left image is of completed steps 1-3b and the right image is of completed steps 1-5.

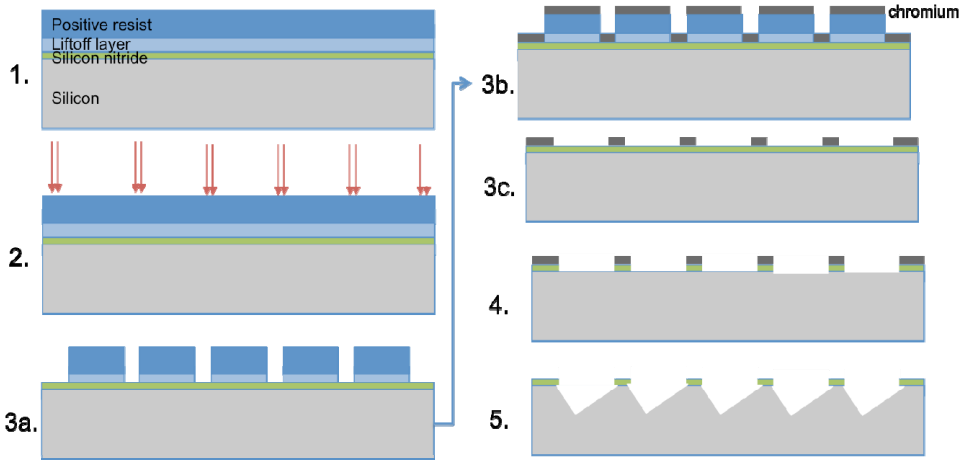


Figure 9. Process flow for chromium liftoff. (1,2,3a) The small lines are written and developed in a bi-layer, positive resist and (3b) chromium is deposited over the surface. (3c) The resist is then dissolved off, leaving small lines of chromium, (4) which then act as the etch mask for removing the silicon nitride. (5) The chromium is removed and silicon nitride acts as the mask for etching the v-grooves.

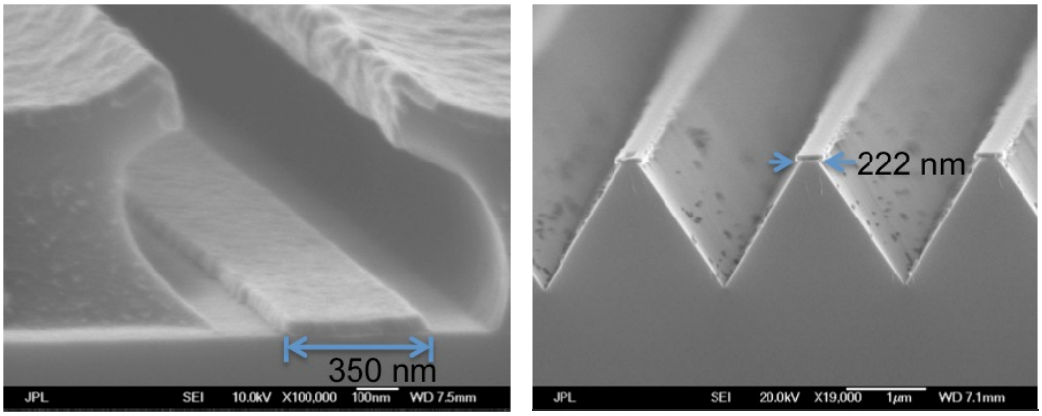


Figure 10. Cross-sectional SEMs of chromium liftoff process on thin test wafer. In the image on the left, steps 1-3b have been completed. The chromium line remaining is 350 nm wide. In the image on the right, steps 1-5 have all been completed. The remaining groove top is now 222 nm because of undercutting in step 5, the KOH groove etch.

3.4 Grating manufacture and testing

After successfully developing and testing the chromium liftoff process on thin silicon wafers, we were then ready to write and process a thick silicon substrate. The completed grating and its front surface interferogram are shown in Figure 11. The total peak-to-valley surface error is 32 nm for this 18° blazed grating, which corresponds to a line placement error of 103 nm.

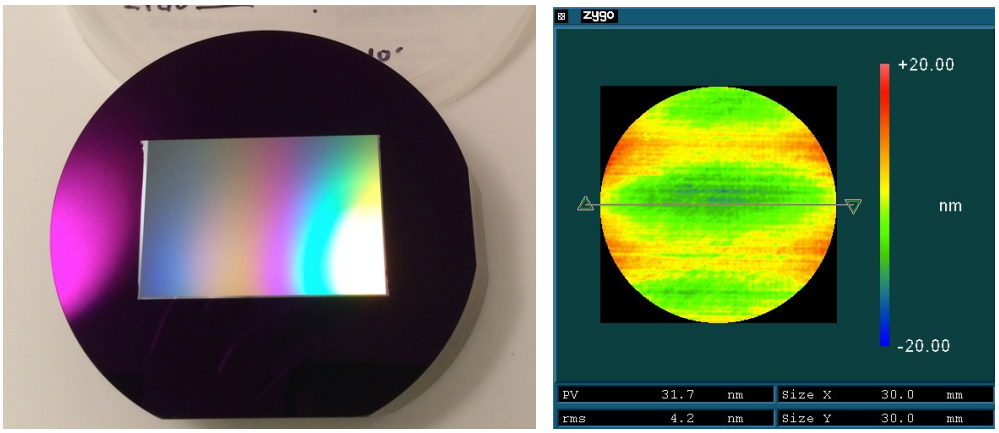


Figure 11. (Left side) Photograph of part K04 patterned by e-beam and chromium liftoff. (Right side) Front surface interferogram (633 nm) in Littrow at the blaze angle of 18° for a 30 mm beam for the same grating.

The monochromatic spectrum of the grating measured in Littrow using a 633 nm laser is shown in Figure 12. The grating has a ghost level $<10^{-4}$. For small periodic errors, the intensity of the Rowland ghosts relative to the main peak is related to the periodic error amplitude by:

$$\frac{I_{ghost}}{I_{line}} = \left[\frac{2\pi n}{\lambda} A \sin\delta\right]^2$$

where A is the amplitude of the spacing error, n is the diffraction order, λ the wavelength, and I_{ghost} and I_{line} are the intensities of the ghost pair and the parent line, respectively. These ghosts correspond to a periodic spacing error of A ≤ 3.3 nm. It should be noted that we have used the method described by Wilson [6] and Gully-Santiago [5] in which the

field size was broken up into several different field sizes to distribute the power, and thus reducing the amplitude, of individual ghosts.

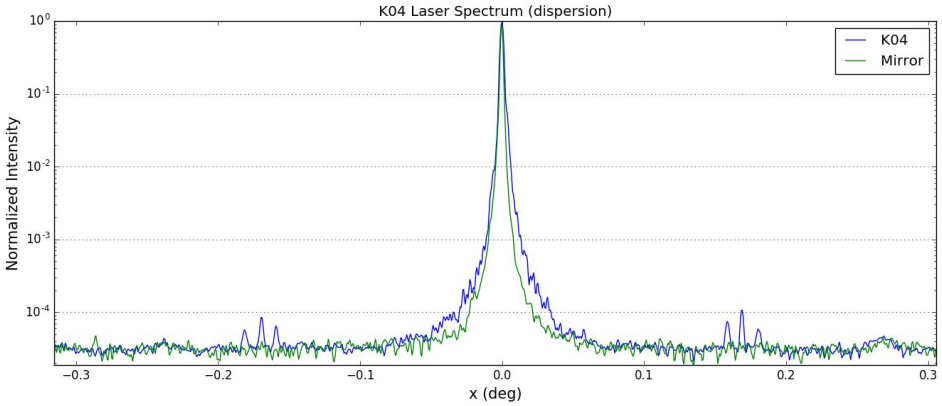


Figure 12. Front-surface monochromatic spectrum of an immersion grating blazed at 18° using a 633 nm HeNe laser. Ghosts below the 10^{-4} level are seen at 0.15°-0.2°, which corresponds to a periodic spacing error of ≤3.3 nm. Several hundred exposures were accumulated to improve the dynamic range of the CCD detector.

4. DIRECT LASER PATTERNING

In Section 2 we described the two-step photolithography process that we developed at the University of Texas for the fabrication of immersed silicon gratings. A photomask is first produced through laser patterning by a commercial photomask vendor. The mask pattern is then transferred onto the silicon grating substrate using contact lithography. The final grating thus contains errors that result from errors in both the mask fabrication and the contact lithography step. It is desirable to eliminate one of the lithography steps and write the grating patterns directly onto the grating substrates to make it easier to meet the stringent pattern placement and line edge roughness requirements outlined in Section 1.

One direct-write photolithography system that is configured for writing on thick, non-standard substrates is the Lumarray ZP-150B parallel laser-beam lithography system [7] operated by the Nano-Structured Optics and Surface Metrology Project at the National Institute of Standards and Technology (NIST). This lithography tool incorporates several design innovations that set it apart from the established laser lithography tools that are used for the fabrication of photomasks and flat panel displays. Four design features are particularly relevant. The ZP-150 tool uses an array of identical zone plates to focus laser light into a photoresist, which gives the technology its name “Zone plate array lithography” (ZPAL). Zone plates can focus light with high numerical aperture and low aberrations resulting in a small focus, enabling precise patterning of relatively small features. Using multiple beams increases the write speed of the lithography tool, which makes the patterning of large areas feasible. Another important feature of the ZP-150 lithography tool is the displacement metrology for the translation stages, which is based on a nano-grid encoder. This metrology system has a high repeatability and facilitates compensation of patterning errors at either the pattern layout stage or the write stage. The ZP-150B lithography tool at NIST was modified to enable the fabrication of diffractive optics on thick, stable, optical glass substrates [8]. It is now one of few direct-write laser lithography systems capable of patterning thick substrates that do not conform to either standard wafer or photomask dimensions.

When patterning silicon wafers with the ZP-150B lithography system we have found that it is always beneficial to use a bottom anti-reflection (BARC) coating between the silicon wafer and the photoresist layer. The exposure wavelength of 403 nm is close to a peak in the refractive index in silicon, which results in a large fraction of light being reflected back into the photoresist. In this section we describe the result of exposure tests that we made on silicon wafers. We tested two commercially available anti-reflection coatings that are suitable for our exposure wavelength of 403 nm. One is the BARLi-II coating already mentioned in Section 2, the other the WiDE-C developer soluble BARC coating by Brewer Science. The two anti-reflective coatings behave differently during the develop process. The WiDE-C coating is soluble in alkaline developer after exposure and thus does not require the additional plasma etch step needed with BARLi-II.

The presence of a nitride layer on top of the silicon substrate complicates the design of the resist material stack because it alters the optical performance of the BARC coatings. The material stack is similar to the one shown in Figure 2. All exposures made with the ZP-150B tool used a positive photoresist layer (Sumitomo PFI-88) with a thickness of 300 nm. The photoresist was spin-coated on a BARC layer (either BARLi-II or WiDE-C), which, in turn, was spun onto the silicon nitride coated wafers. We measured the refractive indices and absorption coefficients of all resist materials using ellipsometry and calculated the reflectance into the photoresist layer as a function of the thicknesses of the silicon nitride and BARC layers. The reflectance calculations were made with a toolbox for Matlab and Octave for calculating properties of optical interference coatings [9]. Figure 13 shows the reflectance into the photoresist with a BARC anti-reflection layer (left) and a WiDE-C anti-reflection layer (right). Since the photoresist layer is much thinner at 300 nm than the resist layer thickness for the contact lithography we decided to reduce the thickness of the silicon nitride layer to make sure the nitride layer can be etched through to the silicon substrate using reactive ion etching (RIE) with the photoresist as an etch mask. We conducted etch tests with silicon nitride coated wafers in KOH and determined that the silicon nitride layer must be at least 50 nm thick to be certain that it is free of pin-holes. For a 50 nm thick silicon nitride layer a minimum of the reflectance into the photoresist is achieved with a 140 nm thick BARLi-II layer (see left side of Figure 13). The developer soluble WiDE-C material requires a layer thickness of about 220 nm as can be seen in Figure 13 on the right. While the thickness of the silicon nitride layer can be controlled very tightly to within ± 1 nm, it can be difficult to achieve reproducible layer thicknesses for spin-coating processes in a research environment when only very small substrate batches are coated. We found that we are generally able to obtain a desired resist material thicknesses to within ± 5 nm if we are careful not to deviate from an established coating process for a resist material e.g. by changing spin times or speeds, or the dispense volume. The reflectance plots in Figure 13 show that a thickness error of ± 5 nm can be tolerated. The relatively large thickness of the WiDE-C layer in relation to the resist thickness made this material more difficult to process because the exposure latitude is reduced. For a fixed development time the resist on a WiDE-C BARC layer requires a higher exposure dose than on a BARLi-II layer because sufficient time is required to fully dissolve the BARC layer during the development step, but this also makes it more likely that the photoresist receives too much dose.

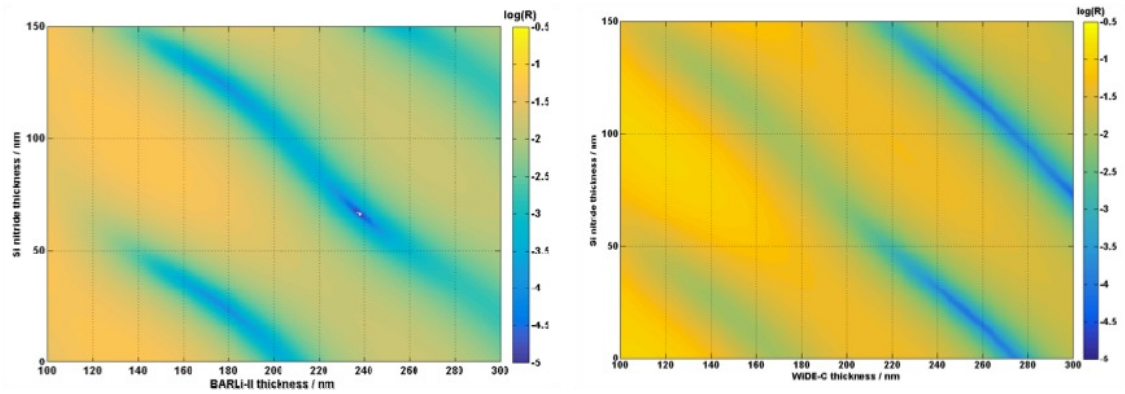


Figure 13: Calculated decadic logarithm of the reflectance into the photoresist at 405 nm with BARLi-II anti-reflection coating (left) and WiDE-C anti-reflection coating (right).

We have patterned thin silicon test wafers with test gratings using the Lumarry ZP-150B system to evaluate the viability of the full fabrication process for gratings patterned with the ZP-150B. In these tests we were comparing the effectiveness of both types of anti-reflective coatings: the BARLi-II used previously in our contact printing experiments and WiDE-C. Figure 14 shows a microscope image of a BARLi-II test wafer after patterning and a scanning electron micrograph of a cross-section of the same wafer after plasma and KOH etching. We have evidence that low-intensity Rowland ghosts close to the main peak occur and are caused by low-spatial-frequency placement errors in the ZP-150B lithography system. We are currently working to reduce the lithography system placement errors, which will reduce the intensity of the ghosts.

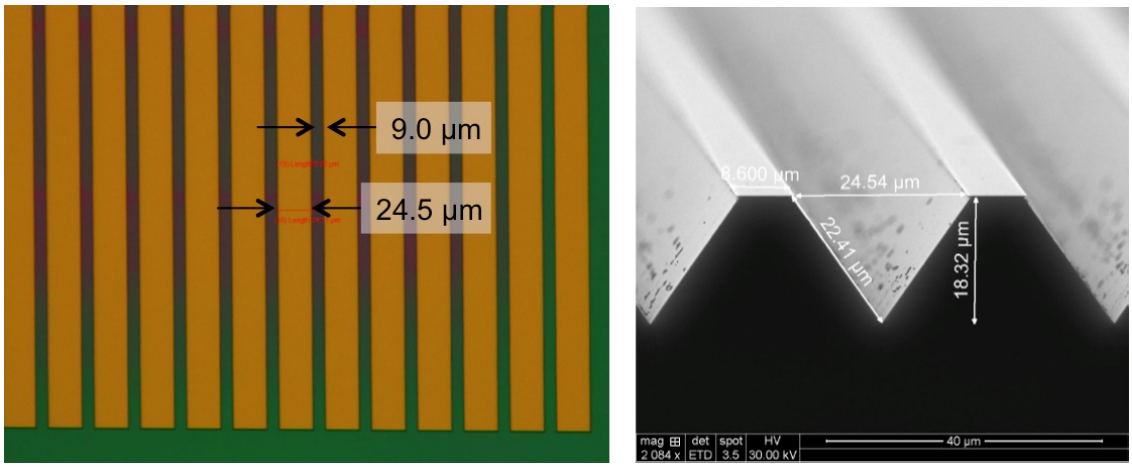


Figure 14. Silicon wafer patterned with a BARLi-II anti-reflective coating under photoresist using the Lumarray ZP-150 after patterning (left) and after etching (right).

5. SUMMARY AND CONCLUSIONS

In this paper we have described several methods of patterning the lines used to etch grooves to form an immersion grating. Our standard process of contact printing was modified to include an anti-reflective coating beneath the photoresist to reduce the contrast of the contacting fringes. The result was a reduction in the amplitude of the fringe component to the phase error by a factor of 1.5. We continued our development of direct electron beam lithography, improving the process by employing a chromium liftoff method that allows us to radically reduce write time and still use a positive electron beam resist. This resulted in a grating surface with excellent phase uniformity and acceptable ghost levels. Finally, we have begun exploring direct laser write lithography, which also allows us to bypass the contact printing step.

As we move forward, we will employ the patterning method most suitable for the type of silicon grating being manufactured. The addition of anti-reflective coatings reduces the effect of contacting fringes, but that still leaves contact printing susceptible to global exposure dose non-uniformity. To that end, we are currently constructing a new UV exposure system that has improved contacting control as well as better UV uniformity over a larger (150 mm x 150 mm) area. The improved contacting control will allow us to better control the parallelism between the photomask and the substrate, allowing us to minimize the fringe effect. For gratings operating in very low order, the grating constants, and thus the line widths, are very small and electron beam patterning is most suitable. However, the electron beam patterning is more prone to Rowland ghosts and may be less suitable for higher blaze angle gratings, which are more sensitive to periodic errors. We will continue our development of direct laser written gratings, as that allows us to bypass the contacting process yet write the course gratings suitable for operation in very high order.

ACKNOWLEDGEMENTS

This work was supported by NASA APRA grant NNX15AD97G and NASA ACT grant NNX12AC31G.

Disclaimer: The full description of the procedures used in this paper requires the identification of certain commercial products and their suppliers. The inclusion of such information should in no way be construed as indicating that such products or suppliers are endorsed by NIST or JPL, or are recommended by NIST or JPL, or that they are necessarily the best materials or suppliers for the purposes described.

REFERENCES

[1] J. P. Marsh, D. J. Mar, and D. T. Jaffe, “Production and evaluation of silicon immersion gratings for infrared astronomy," Applied Optics 46, pp. 3400-3416, June 2007.

[2] Weisong Wang, Michael Gully-Santiago, Casey Deen, Douglas J. Mar and Daniel T. Jaffe, "Manufacturing of silicon immersion gratings for infrared spectrometers", *Proc. SPIE* 7739, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation, 77394L (July 20, 2010)

[3] Cynthia B. Brooks, Michael Gully-Santiago, Michelle Grigas, Daniel T. Jaffe, "New metrology techniques improve the production of silicon diffractive optics", Proceedings of SPIE Vol. 9151, 91511G (2014)

[4] Shipley Microposit® S1800® Series Photo Resists product literature. http://www.microchem.com/PDFs_Dow/S1800.pdf

[5] Michael A. Gully-Santiago, Daniel T. Jaffe, Cynthia B. Brooks, Daniel W. Wilson, Richard E. Muller, "High performance Si immersion gratings patterned with electron beam lithography", Proceedings of SPIE Vol. 9151, 91515K (2014)

[6] Daniel W. Wilson, Richard E. Muller, Pierre M. Echternach, and Johan P. Backlund, “Electron-beam lithography for micro and nano-optical applications,” Proceedings of SPIE Vol. 5720, p. 68-77 (2005)

[7] H. I. Smith, M. E. Walsh, F. Zhang, J. Ferrera, G. Hourihan, D. Smith, R. Light, and M. Jaspan, “An innovative tool for fabricating computer-generated holograms,” J. Phys. Conf. Ser. 415, 012037 (2013).

[8] Q. Wang and U. Griesmann, “Versatile bilayer resist for laser lithography at 405 nm on glass substrates”, Opt. Eng. 52, 105104 (2013)

[9] Thin Film Toolbox: https://sites.google.com/site/ulfgri/numerical/thin-films

Brooks, Cynthia; Griesmann, Ulf; Grigas, Michelle; Jaffe, Daniel; Kidder, Benjamin; Muller, Richard; Wilson, Daniel.
”Process improvements in the production of silicon immersion gratings.”
Paper presented at Astronomical Telescopes and Instrumentation, Edinburgh, United Kingdom. June 25, 2016 - July 1, 2016.

STUDY ON SPECTRAL CHARACTERISTICS FOR IDENTIFICATION OF SKIN COLOUR OF INJURED JAPANESE PERSONS AT DISASTER SITES

Akizuki, Y.[1], Ohno, Y.[2]
[1] University of Toyama, Toyama, JAPAN,
[2] National Institute of Standards and Technology, Gaithersburg, MD USA
akizuki@edu.u-toyama.ac.jp

Abstract

In order to develop suitable light sources for disaster medical work, we first determined the skin colour and spectral reflectance of injured or sick persons under shocked or congested states. Using the NIST Color Quality Scale (CQS) simulation model, we theoretically determined appropriate spectral power distributions (SPD) of a light source which is effective in distinguishing colour differences among various skin states. Moreover, in order to evaluate the colour rendering, a subjective qualitative evaluation was conducted with lights of similar SPDs to the theoretical SPDs in the NIST Spectrally Tunable Lighting Facility.

Keywords: colour rendering, skin colour, disaster medical work, spectral characteristics

1 Introduction

There are many disasters in Japan every year. It is widely known that the critical response time for saving injured persons alive is within 72 hours, so the rescue and medical teams operate for entire days and nights. The limited performance of light sources used for rescue work makes it more difficult for rescue workers to observe situations, especially at night and in confined spaces. During relief efforts in confined spaces after earthquake disasters, it is important to diagnose injured persons as a case of crush syndrome or not, based on their skin colour.

Light emitting diode (LED) lights are increasingly used for such rescue work, but some medical relief workers are concerned that LED lights make it difficult to observe an injured person's skin colour, as they have distinct spectral power distributions (SPD) dissimilar to incandescent or florescent lamps. In our previous research, we used a questionnaire to survey visual problems experienced by disaster relief workers. Many rescuers and medical staff members indicated that major problems of personal lighting equipment were due to the lack of luminous flux, narrow spatial distribution of light, and poor colour rendering (Akizuki et al. 2011). These results revealed that relief workers were well aware of those serious visual and lighting problems, as they work on the front line of disaster relief efforts.

In order to develop suitable light sources for disaster medical work, it is first necessary to determine the skin colour of injured or sick persons under shocked or congested states. Next, we need to determine appropriate SPDs for light sources, which are effective at distinguishing colour differences among various skin states. Our research goal is to create more effective lighting equipment and environments in disaster sites. In this paper, we propose a basic method of analysis to determine the spectral characteristics of light sources for the identification of injured or sick persons' skin colour in a qualitative experiment.

2 Spectral Reflectance of Japanese Injured or Sick Persons' Skin Colour

In the preceding research, we carried out an experiment for collecting skin's spectral reflectance data under artificially-produced shocked or congested state of healthy subjects (Akizuki et al. 2013, 2016). The skin colour of critical patients in shock conditions changes widely over time. Therefore, we artificially produced the injured skin colour, which was the most prominent symptom of shock from distal ischemia in healthy subjects. The previous study is summarized below.

Subjects consisted of Japanese students at the University of Toyama: 22 young female subjects with an average age of 22 years old and 15 young male subjects with an average age of 20

Akizuki, Yuki; Ohno, Yoshihiro.
"Study on Spectral Characteristics for Identification of Skin Colour of Injured Japanese Persons at Disasters Sites."
Paper presented at CIE Tutorial and 4th CIE Symposium on Colour and Visual Appearance, Prague, Czech Republic. September 5, 2016 - September 7, 2016.

years old. Moreover, registered members of the Toyama Silver Human Resources Center were recruited as additional experimental subjects: 15 elderly female subjects with average age of 69 years old and 15 elderly male subjects with an average age of 72 years old. Their blood pressure, heart rate, and the percentage saturation of oxygen were measured to check for health conditions before the experiments. All subjects wore a thin shirt with long sleeves during the experiments.

The experimental timetable is shown in Figure 1. Before each experimental session, a subject sat on a chair quietly for 60 s. Then, the subject put his or her hand on a table and the skin colour of the back of the hand between the first metacarpal and the second metacarpal was measured as the "healthy skin state" by a spectrophotometer. After the experiment started (t=0 s), the subject raised his or her upper arm, held it in that position for 90 s in order to reduce the flow of blood to peripheral portions, such as the hand. The upper arm was wrapped with a sphygmomanometer cuff (UM-101, AND). After 90 s of maintaining the position (t=90 s), the upper arm was brought to 200 mmHg pressure by the sphygmomanometer cuff, and the pressure was maintained for 60 s. Internal bleeding was not observed. After that (t=150 s), the subject slowly pulled his or her arm down on the table, and remained in that position for 60 s. Then (t=210 s), the skin colour of the same portion of the hand was measured as the "ischemia-shocked skin state" by the spectrophotometer. When a total of 240 s elapsed from the beginning of the experiment (t=240 s), the pressure of the sphygmomanometer cuff was reduced to 0 mmHg, and the skin colour of same part of the hand was measured as "the reperfusion-congested skin state" in 10 s intervals for 240 s. Both hands of a subject were measured on separate trails. At first, the right hand was measured for three skin states and then the left hand was measured. In this experiment, we measured spectral reflectance of three skin states: healthy skin, shocked skin, and congested skin. By using these spectral reflectance data, we calculated XYZ tristimulus values under the standard illuminant D65. We defined the spectral reflectance data which showed the lowest Y value between t=240 s to t=480 s as "the congested skin state" spectral reflectance. The time at which the skin exhibited this "the congested skin state" varied between individuals. Figure 2 shows the experimental setup.

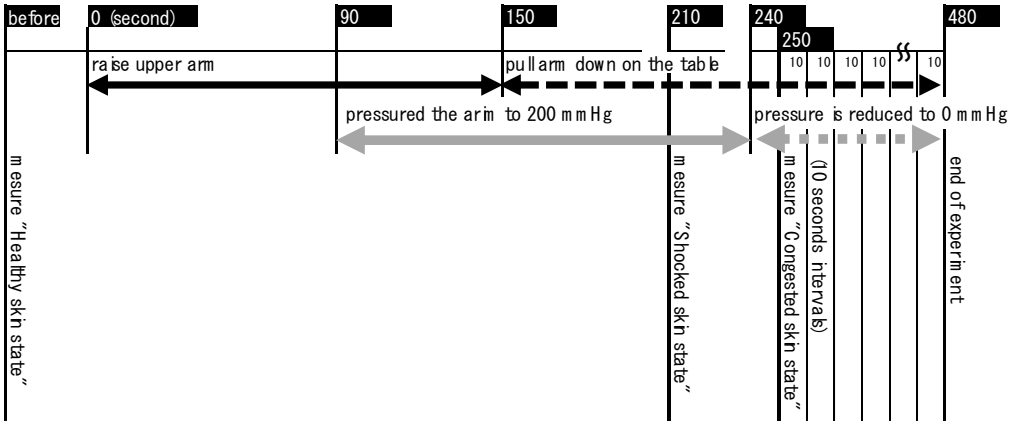


Figure 1 – Experimental Timetable for Measuring Injured Skin Colour

(1) Starting Position (2) Measurement Position (3) Skin Colour Example

Figure 2 – Experimental Setup

When compared with the healthy skin, the shocked skin tended to be more yellowish and to have higher lightness, and the congested skin tended to be more reddish and to have higher saturation and lower lightness. These tendencies were found in most subjects. Therefore the fundamental principle in the selection process for the skin colour database of this research was that the order of Y values was “Shocked state”>”Healthy state”>”Congested state”. If the order of Y values was not same as this general finding, the result was discarded. We collected more than 22 skin states’ data for each age and gender group.

Since the values of Munsell Colour System obtained by the spectrophotometer under the standard illuminant D65 were reported by the instrument as decimal numbers, we rounded these off to integer values. With consideration for parameters of the commercially available skin colour charts made by Japan Color Enterprise Co. Ltd., such as “the series of Skin Colour 75”[1], we set up the Munsell Hue 2,5 interval, Munsell Value 1.0 interval, and Munsell Chroma 1,0 interval for our skin colour chart of injured or sick Japanese persons. We categorized all the results by three skin states (healthy, shocked, and congested skin) and subject groups. Typical results are shown in Table 1. There were significant differences between age, gender and skin state groups ($p<0,05$).

Table 1 – Typical Skin Colour of Injured or Sick Japanese Persons according to Age and Gender (illuminated by standard light source D65)

	Healthy Skin		Shocked Skin		Congested Skin	
	Munsell	ratio	Munsell	ratio	Munsell	ratio
Young Female	7,5YR6/3	47% (=17/36)	10YR7/3	53% (=19/36)	5YR6/4	44% (=14/32)
Young Male	7,5YR6/3	26% (= 6/23)	10YR6/4	52% (=12/23)	5YR6/4	41% (= 9/22)
Elderly Female	7,5YR6/3	63% (=17/27)	10YR6/3	57% (=16/28)	5YR6/4	32% (= 9/28)
Elderly Male	5YR5/4	42% (=10/24)	10YR6/3	33% (= 8/24)	5YR5/4	44% (=10/23)

Next, we selected the typical subject spectral skin colour for each of the four subject groups. The results are shown in Figure 3. They had the same skin colour of Munsell values as Table 1. According to a typical result of the young female subjects in Figure 3 (upper left), the spectral reflectance of the shocked skin was whitened or higher than healthy skin, especially within the range from 500 nm to 600 nm. On the other hand, the spectral reflectance of the congested skin was lower than other skin states. There were significant differences of spectral reflectance between 500 nm to 600 nm in all groups.

[1] Specific firms and trade names are identified in this paper to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Akizuki, Yuki; Ohno, Yoshihiro.
”Study on Spectral Characteristics for Identification of Skin Colour of Injured Japanese Persons at Disasters Sites.”
Paper presented at CIE Tutorial and 4th CIE Symposium on Colour and Visual Appearance, Prague, Czech Republic. September 5, 2016 - September 7, 2016.

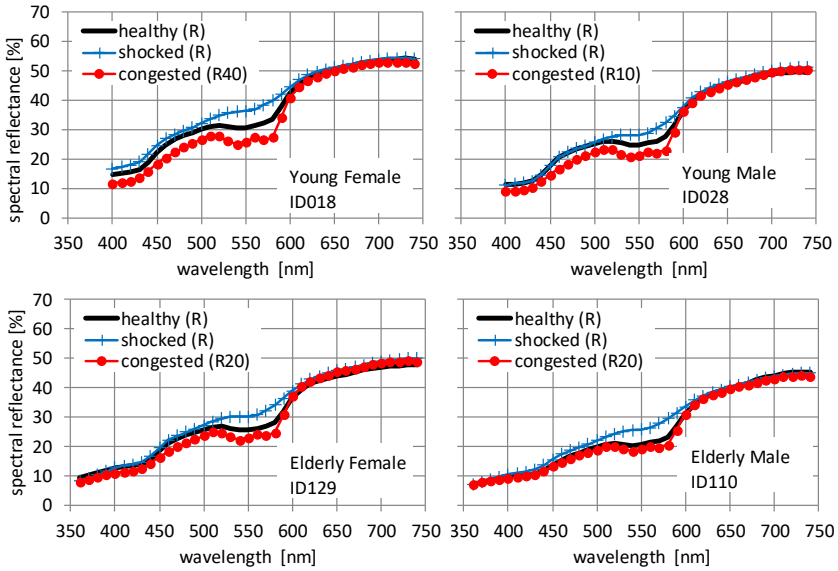


Figure 3 – Spectral Reflectance of Typical Subjects' Skin Colour

Unlike the real skin's spectral reflectance distribution, the commercially available skin colour charts made by Japan Color Enterprise Co. Ltd. has spectral reflectance distributions with lower reflectances than real skin colours at 600 nm and above, even with same tristimulus values (YXZ) (see Figure 4). Therefore we should use the real skin data of Figure 3, instead of the skin colour charts, in order to determine SPDs of light sources for identification of skin colour of injured or sick Japanese persons.

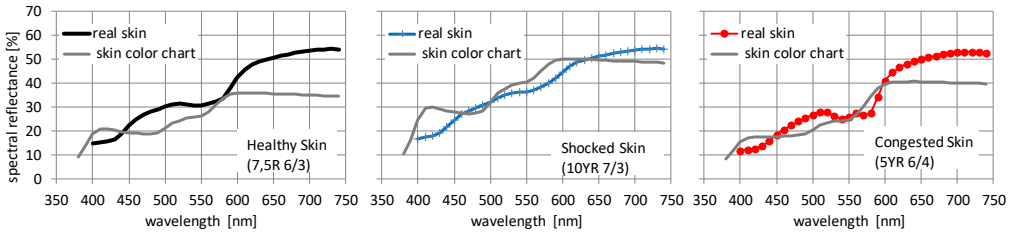


Figure 4 –Comparison of Spectral Refelctance Between Real Skin and Skin Colour Chart (Young Female Subject ID018)

3 Theoretical Spectral Power Distribution for Identification of Skin Colour

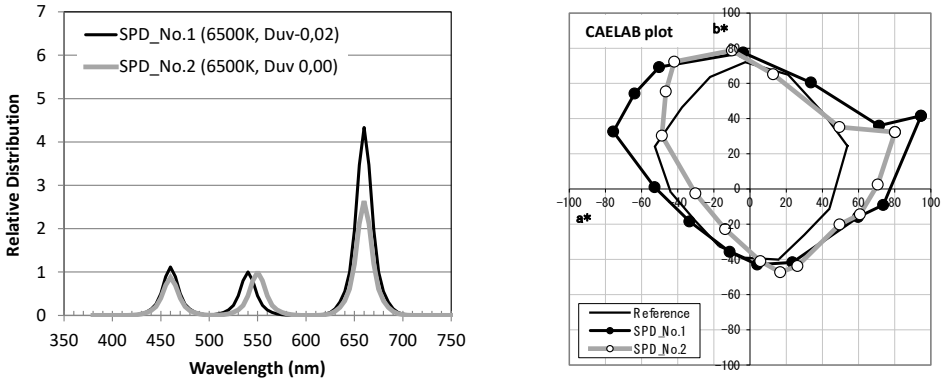
3.1 Simulation Procedures

In order to determine a theoretical ideal light source SPD, which produces the maximum colour difference ΔE^*_{ab} between the healthy skin and the shocked skin or congested skin, we used a spreadsheet program developed for Color Quality Scale (CQS) (Davis and Ohno 2010). This program provides calculation results of colour quantities, such as the CIE Colour Rendering Index (CRI) and the values of CQS for given light source SPDs. Moreover, this program can simulate a white LED SPD made of three-LED or four-LED chips. On both simulation models, with the minimization tool in the spreadsheet, we can determine various optimized LED SPDs just by entering the values of target correlated colour temperature (CCT), target (ANSI, 2015), the LEDs' peak wavelengths and spectral width (full width at half maximum), and so on.

In this research, we fixed the LEDs' spectral width at 20 nm. Target CCT was set at 6500 K and 2700 K. Colour differences were measured for LEDs' peak wavelengths at every 10 nm within the range from 450 nm to 660 nm and for Duv at every 0,01 within the range of -0,02 to +0,02. We used the real skin data shown in Figure 3 to calculate the colour difference ΔE^*_{ab} between the healthy skin and the shocked skin or the congested skin per person.

3.2 Results based on the White LED Simulation Model by three-LED chips

By using the white LED simulation model for three-LED chips, we determined the theoretical SPDs that produced the maximum ΔE^*_{ab} at a CCT of 6500 K. The results are shown in Figure 5, Table 2 and Table 3. The right image in Figure 5 shows a CIE 1976 $L^*a^*b^*$ plot of colour rendering performance with the 15 saturated reflective samples in the Colour Quality Scale (CQS) (Davis and Ohno 2010). Table 2 shows two indices of the CIE Colour Rendering Index (average CRI R_a and strong red's CRI R_9), and three CQS scales (general colour quality scale Q_a, colour fidelity scale Q_f, and gamut area scale Q_g). The spectrum "SPD_No.1" produced the maximum colour difference ΔE^*_{ab} between the healthy skin and the shocked skin (ΔH-S) or the congested skin (ΔH-C) in most combinations. Compared with the reference illuminant (blackbody at the same CCT), "SPD_No.1" produced more than 1,42 times ΔE^*_{ab}, and the average ratio of "SPD_No.1" and the reflectance was 1,59. However the colour rendering results of "SPD_No.1" were very poor, especially CRI R_9. Therefore, we optimized the theoretical SPDs again, the results of which had better colour rendering (CRI R_a>50) and produced a larger ΔE^*_{ab} ("SPD_No.2"), except for young male subject's ΔH-S, whose colour differences ΔE^*_{ab} decreased slightly. However, the colour rendering was improved significantly.



(a) Spectrum with three-LED chips **(b) Appearance of the CQS samples**

Figure 5 – Theoretical SPDs that produced the maximum ΔE^*_{ab} (6500 K, 3-LED)

Table 2 – Theoretical SPDs' Characteristics (6500K, 3-LED)

	peak wavelength (width)	CCT	Duv	CRI Ra	CRI R9	CQS Qa	CQS Qf	CQS Qg
SPD_No.1	460 (20) 540 (20) 660 (20)	6500	-0,02	10	-425	27	27	157
SPD_No.2	460 (20) 550 (20) 660 (20)	6500	0,00	53	-201	73	71	120

Table 3 – Skin Colour Differences ΔE^*_{ab} with Theoretical SPDs (6500K, 3-LED)

	Young Female		Young Male		Elderly Female		Elderly Male	
	ΔH-S	ΔH-C	ΔH-S	ΔH-C	ΔH-S	ΔH-C	ΔH-S	ΔH-C
SPD_No.1	8,55	8,69	6,49	8,16	7,51	7,30	10,86	3,26
SPD_No.2	7,91	6,80	6,64	6,38	7,31	5,46	9,74	2,75
Reference	4,78	5,08	4,67	5,10	5,30	4,22	6,57	2,30

Akizuki, Yuki; Ohno, Yoshihiro.
"Study on Spectral Characteristics for Identification of Skin Colour of Injured Japanese Persons at Disasters Sites."
Paper presented at CIE Tutorial and 4th CIE Symposium on Colour and Visual Appearance, Prague, Czech Republic. September 5, 2016 - September 7, 2016.

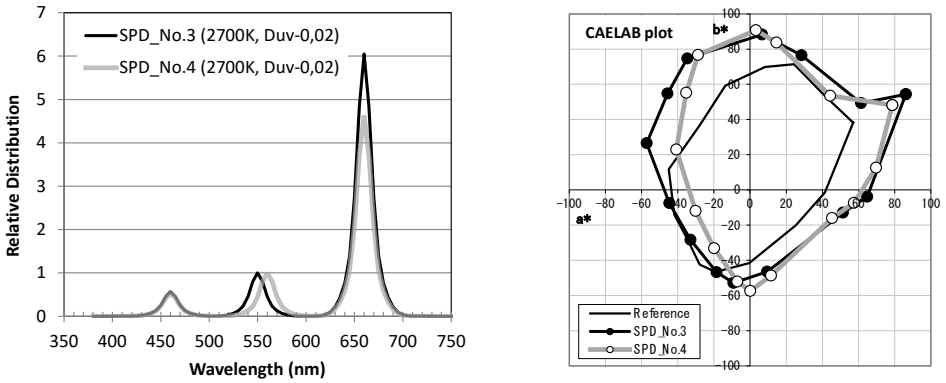
Next, we optimized the theoretical SPDs that produced the maximum ΔE^*_{ab} at a CCT of 2700 K. The results are shown in Figure 6, Table 4 and Table 5. Compared with the reference illuminant, "SPD_No.3" produced more than 1,26 times ΔE^*_{ab} and the average ratio of "SPD_No.3" and the reflectance was 1,40. However the CRI values of "SPD_No.3" was very poor. Therefore we optimized again, generating theoretical SPDs which had better colour rendering (CRI R_a>50) and produced larger ΔE^*_{ab} values ("SPD_No.4"), except for the ΔH-S of young male and elderly female subjects, whose colour differences ΔE^*_{ab} decreased slightly. However, the colour rendering was improved, especially CRI R_a and R_9.

Table 4 – Theoretical SPDs' Illuminant Characteristics (2700K, 3-LED)

	peak wevelength (width)	CCT	Duv	CRIRa	CRIR9	CQS_Qa	CQS_Qf	CQS_Qg
SPD_No.3	460(20) 550(20) 660(20)	2700	−0,02	19	−286	72	64	172
SPD_No.4	460(20) 560(20) 660(20)	2700	−0,02	54	−180	73	70	137

Table 5 – Skin Colour Differences ΔE^*_{ab} with Theoretical SPDs (2700K, 3-LED)

	Young Female		Young Male		Elderly Female		Elderly Male	
	ΔH-S	ΔH-C	ΔH-S	ΔH-C	ΔH-S	ΔH-C	ΔH-S	ΔH-C
SPD_No.3	7,93	7,83	6,05	7,55	6,80	6,63	9,57	2,63
SPD_No.4	7,49	6,82	6,29	6,64	6,83	5,71	9,03	2,44
Reference	5,02	5,48	4,57	5,53	5,06	4,72	6,42	2,09



(a) Spectrum with three-LED chips **(b) Appearance of the CQS samples**

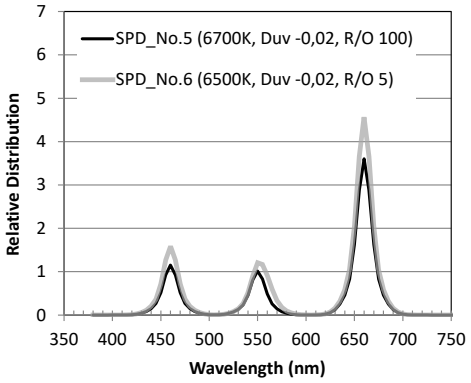
Figure 6 – Theoretical SPDs that produced the maximum ΔE^*_{ab} (2700 K, 3-LED)

3.3 Results based on the White LED Simulation Model by four-LED chips

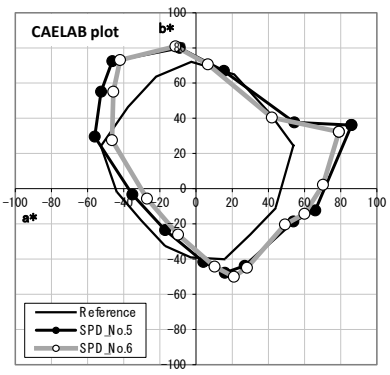
Overall, the colour differences ΔE^*_{ab} at the CCT of 2700 K were lower than the ΔE^*_{ab} at the CCT of 6500 K. However, the colour rendering results at 2700 K were better than the ones at 6500 K. Moreover, it would appear that the spectral range between 540 nm to 560 nm might impact both the identification of injured or sick skin colour and the colour rendering.

Therefore, by using the white LED simulation model for four-LED chips, we optimized theoretical SPDs again, which produced the maximum ΔE^*_{ab} between healthy skin and shocked skin (ΔH-S) or congested skin (ΔH-C) at the CCT of 6500 K. For this simulation, we also had to set the spectral power ratio of the red LED to the orange LED (red to orange ratio). The results are shown in Figure 7, Table 6 and Table 7. Note that, in Figure 7, two LED peaks overlap and appear as three peaks. "SPD_No.5" produced the maximum colour difference ΔE*ab in this simulation. Compared with the reference illuminant, "SPD_No.5" produced more than 1,26 times

ΔE^*_{ab} and the average ratio of "SPD_No.5" and the reflectance is 1,46. Another theoretical SPD, which had better colour rendering (CRI R_a>50) and produced larger ΔE^*_{ab}, was derived ("SPD_No.6"). "SPD_No.6" produced the maximum ΔE^*_{ab}, while also having good colour rendering, of all theoretical SPDs generated.



(a) Spectrum with four-LED chips



(b) Appearance of the CQS samples

Figure 7 – Theoretical SPDs that produced the maximum ΔE^*_{ab} (6500 K, 4-LED)

Table 6 – Theoretical SPDs' Characteristics (6500K, 4-LED)

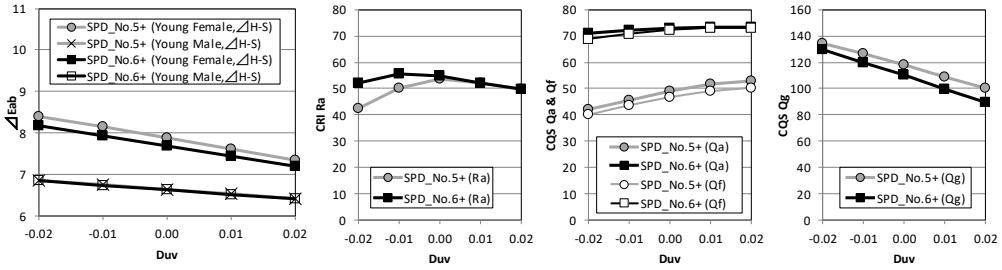
	peak wavelength (width)	R/O	CCT	Duv	CRIRa	CRIR9	CQS_Qa	CQS_Qf	CQS_Qg
SPD_No.5	460(20) 550(20) 560(20) 660(20)	100	6500	−0,02	42	−306	70	66	141
SPD_No.6	460(20) 550(20) 560(20) 660(20)	10	6500	−0,02	52	−260	71	69	130

Table 7 – Skin Colour Differences ΔE^*_{ab} with Theoretical SPDs (6500K, 4-LED)

	Young Female		Young Male		Elderly Female		Elderly Male	
	ΔH-S	ΔH-C	ΔH-S	ΔH-C	ΔH-S	ΔH-C	ΔH-S	ΔH-C
SPD_No.5	8,41	7,32	6,84	6,87	7,59	5,98	10,34	2,89
SPD_No.6	8,18	6,91	6,84	6,49	7,52	5,60	10,05	2,80
Reference	4,78	5,08	4,67	5,10	5,30	4,22	6,57	2,30

3.4 Relationship among Colour Difference, Colour Rendering and Duv

The relationship among ΔE^*_{ab} between the healthy skin and the shocked skin (ΔH-S), Duv and colour rendering results are shown in Figure 8. The peak wavelength, spectral width and the red to orange ratio were not changed from the conditions of "SPD_No.5" and "SPD_No.6", and only Duv values were changed for this simulation. Regardless of subject group, ΔE^*_{ab} decreased when Duv increased. The CQS results (Q_a, Q_f and Q_g) show a linear relationship with Duv. CRI R_a does not show a linear relationship with Duv, and the Duv value reaches a peak depending on the SPD. Future light source development for disaster medical works should take into consideration both the identification of diseased skin colour and the colour rendering of the surrounding environment. In any case, the results show the largest ΔE^*_{ab} at Duv=-0,02. While visual experimental data on Duv for general conditions are available (Ohno and Fein, 2014), the authors were concerned that such large shift in Duv with a very saturated color gamut may not appear natural for the surrounding environment, as well as for real skin colors. Thus, we conducted a preliminary visual evaluation experiment related to Duv shift, as reported in the next section.



(1) ΔE*ab of two subject groups (2) CRI Ra (3) CQS Qa & Qf (4) CQS Qg

Figure 8 – Relationship among ΔE^*_{ab}, Colour Rendering and Duv

4 Evaluative Experiment

To evaluate the acceptability of the Duv magnitude and perceived color quality of the theoretical SPD presented in the previous section, an attempt was made to simulate the SPD on NIST Spectral Tunable Lighting Facility (STLF), and visual evaluations on the interior room, skin tone, and the skin color charts, were conducted.

4.1 Experimental Procedures

The NIST STLF does not have efficient narrow-band green LEDs at a peak wavelength near 550 nm, as such LEDs are not available in the current market. Therefore, we could not reproduce the theoretical SPDs presented in the previous section. We produced spectra as closely as possible to the theoretical SPD using the STLF's 25 channels, and conducted visual evaluation experiments with one subject (the first author) with normal color vision. Since the spectra are not very close, this is a preliminary study for vision experiments in the future.

The NIST STLF consists of two illuminated rooms with off-white walls. The dimensions of each room are 2,5 m in width, 2,5 m in length, and 2,5 m in height. Both rooms have 25 channels of LED spectra with peaks from 450 nm to 650 nm. Spectral distribution, CCT, Duv and illuminance can be controlled independently. In this research, we used the right side room. The illuminant conditions are shown in Figure 9. The experiments were conducted with four SPDs and nine different Duv levels from Duv=-0,025 to Duv=0,015. The illumination characteristics (ΔE^*_{ab} between the healthy skin and the shocked skin (ΔH-S), and CRI Ra) are shown in Figure 10. All ΔE^*_{ab} values were lower than the theoretical SPDs in the preceding section. The illuminant "2700K LED" had the lowest colour rendering among the experimental illuminant conditions.

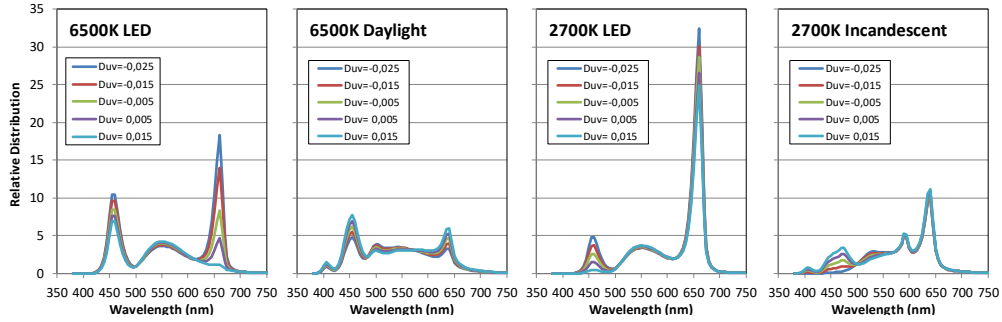
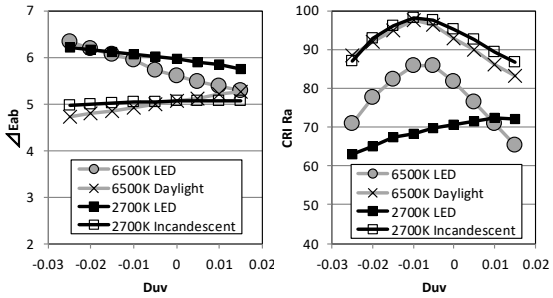


Figure 9 – Spectral Power Distributions in the Evaluative Experiments



(a) ΔE^*_{ab} of young female's H-C **(b) CRI R_a**

Figure 10 – Illumination Characteristics

Figure 11 –Experimental Setup

The experimental setup is shown in Figure 11. This experimental procedures were similar to the previous research (Ohno and Fein 2014). The subject sat on a couch placed at the right side room of NIST STLF. The room has a simulated interior room environment with a table, a bookshelf with some books, artificial flowers, a mirror and two paintings in the room. Some artificial fruits and vegetables were on the table. The illuminance on the table was set to 200 lx. For each combination of CCT and SPD, the subject was first adapted to the illumination at one end of Duv (e.g., Duv=0,01) for five minutes in order to be adapted to the illumination completely. After adaptation, the subject was asked whether the lighting was acceptable or not. Then, a pair of lights, which had a Duv level higher or lower than 0,005 Duv (so in the case, Duv=0,015 and 0,005), was presented. The pair of lights was alternated three times, with each alternation accompanied by a sound from a computer. During the time, the subject was asked three questions; "Which light presents the whole interior more naturally?," "Which light presents your hand's skin colour more naturally?" and "Which light presents the skin colour charts (which had same Munsell values of young female in Table 1) more distinguishable?". Then, the next Duv was presented (e.g. Duv=0), the subject was adapted to the illumination for one minute, and another trial with a pair of lights (Duv=0,005 and -0,005) was made. This was repeated for four levels of adapting Duv (0.01, 0, -0,01 and -0,02), which completed one run of one combination of CCT and SPD. A similar run of the experiment was then conducted for the different combination of CCT and SPD. Then, another run with the first combination of CCT and SPD, but with the reverse order of Duv (start from Duv=-0,02 and ends at Duv=0,01) was conducted. A run for each Duv direction at each combination of CCT and SPD was repeated once. Therefore, there were two runs for each combination of CCT and SPD.

The results are shown in Table 8. In this table, "ascending" means that the presented order of Duv condition was from -0,02 to 0,01, and "descending" means that the order was from 0,01 to -0,02. "Lower Duv" means that the subject chose the lower Duv light (chromaticity shift in pinkish direction) in the pair as more natural, and "higher Duv" meant that the subject chose the higher Duv light (shift in yellowish direction) as more natural. The gray hatching illuminant condition

means the subject had had much difficulty responding. “2700K LED” was not acceptable for any Duv. The subject felt that the whole interior was most natural when illuminated by light of around Duv=-0,015 for most lights. On the other hand, the subject felt that her hand’s skin colour was most natural when illuminated with light around Duv=-0,005, except for “2700K LED”. The results for the skin colour charts were similar to those for the whole interior, and different from those for subject’s real skin colour. This was possibly because the spectral reflectance of the skin chart does not accurately represent real skin (in Figure 4). “6500K LED” produced the maximum ΔE^*_{ab} with good colour rendering.

Table 8 –Raw Results for Subject

(1) Tolerance of the light

	6500K LED		6500K Daylight		2700K LED		2700K Incandescent	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0,01	not	**Accept**	**Accept**	**Accept**	not	not	not	not
0	**Accept**	**Accept**	**Accept**	**Accept**	not	not	**Accept**	**Accept**
-0,01	**Accept**	**Accept**	**Accept**	**Accept**	not	not	**Accept**	**Accept**
-0,02	not	**Accept**	not	not	not	not	**Accept**	**Accept**

(2) Natural-looking whole interior

	6500K LED		6500K Daylight		2700K LED		2700K Incandescent	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0,01	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
0	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
-0,01	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
-0,02	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	lower Duv	lower Duv	**Higher Duv**	**Higher Duv**

(3) Natural-looking hand’s skin color

	6500K LED		6500K Daylight		2700K LED		2700K Incandescent	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0,01	lower Duv	lower Duv	lower Duv	lower Duv	**Higher Duv**	lower Duv	lower Duv	lower Duv
0	lower Duv	lower Duv	lower Duv	lower Duv	**Higher Duv**	**Higher Duv**	lower Duv	lower Duv
-0,01	**Higher Duv**	**Higher Duv**	lower Duv	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**
-0,02	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**

(4) Identification of skin color charts

	6500K LED		6500K Daylight		2700K LED		2700K Incandescent	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0,01	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
0	lower Duv	0 (0)	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
-0,01	lower Duv	**Higher Duv**	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	**Higher Duv**
-0,02	**Higher Duv**	**Higher Duv**	**Higher Duv**	lower Duv	**Higher Duv**	lower Duv	**Higher Duv**	**Higher Duv**

5 Conclusions

We theoretically determined the effective spectral characteristics for identification of skin colour of injured or sick Japanese persons. Highly realistic injured or skin colour samples, which have the same spectral characteristics as human real skin, are needed, as are quantitative subjective experiments, in order to obtain valid and universal results.

Theoretical LED SPDs increase the colour difference ΔE^*_{ab} between healthy skin and shocked skin or congested skin of typical subjects compared to reference light or actual LED light sources. Therefore, the development of narrow-band green LED emitters with peak wavelength around 550 to 560 nm is desired.

Acknowledgments

This research was partly supported by the Research Fellowship for Female Researchers in University of Toyama, and the Grants-in-Aid for Scientific Research of Japanese Society for the Promotion of Science (No.C-16K06607, Program Leader: Yuki Akizuki).

Akizuki, Yuki; Ohno, Yoshihiro.
”Study on Spectral Characteristics for Identification of Skin Colour of Injured Japanese Persons at Disasters Sites.”
Paper presented at CIE Tutorial and 4th CIE Symposium on Colour and Visual Appearance, Prague, Czech Republic. September 5, 2016 - September 7, 2016.

References

AKIZUKI, Y. SUZUKI, H. YOSHIMURA, A. OHYAMA, F. AKITOMI, S. KAKO, Y. 2011. Lighting Equipment Problems of Outdoor Rescue Work and Medical Activity at Night. *Proceedings of 27th session of the CIE,* 779-786.

AKIZUKI, Y. TOMOHIRO, J. WAKASUGI, M. YOSHIMURA, A. TAKASHINA, K. 2013. Creation of Colour chart database of disaster victim's skin - Measurement of quasi-skins in shock and congested state by healthy young subjects -. *Proceedings of the 12th International Colour Association*, 4pages.

AKIZUKI, Y. NAKAGAWA, N. 2016. Study on Japanese Skin Color Chart of Injured or Sick Persons according to age and gender. *Proceedings of AIC2016 Interim Meeting*, 4pages.

ANSI. 2015. ANSI_NEMA_ANSLG, C78.377-˘2015 Specifications for the Chromaticity of Solid State Lighting Products.

CIE 13.3. 1995. Method of measuring and specifying colour rendering properties of light sources.

DAVIS, W. OHNO, Y. 2010. Color quality scale. *Optical Engineering, Vol.49 (3),* pp.033602-1-16.

OHNO, Y. FEIN, M. 2014. Vision Experiment on Acceptable and Preferred White Light Chromaticity for Lighting, *CIE x039*, 192-199.

Akizuki, Yuki; Ohno, Yoshihiro.
"Study on Spectral Characteristics for Identification of Skin Colour of Injured Japanese Persons at Disasters Sites."
Paper presented at CIE Tutorial and 4th CIE Symposium on Colour and Visual Appearance, Prague, Czech Republic. September 5, 2016 - September 7, 2016.

When "Optimal Filtering" Isn't

J. W. Fowler, B. K. Alpert, W. B. Doriese, J. Hays-Wehle, Y.-I. Joe, K. M. Morgan, G. C. O'Neil, C. D. Reintsema, D. R. Schmidt, J. N. Ullom, and D. S. Swetz

***Abstract*—The so-called "optimal filter" analysis of a microcalorimeter's x-ray pulses is statistically optimal only if all pulses have the same shape, regardless of energy. However, the shapes of pulses from a nonlinear detector can and do depend on the pulse energy. A pulse-fitting procedure that we call "tangent filtering" accounts for the energy dependence of the shape and should, therefore, achieve superior energy resolution. We take a geometric view of the pulse-fitting problem and give expressions to predict how much the energy resolution stands to benefit from such a procedure. We also demonstrate the method with a case study of K-line fluorescence from several 3d transition metals. The method improves the resolution from 4.9 to 4.2 eV at the Cu Kα line (8.0 keV).**

***Index Terms*—Superconducting photodetectors, linearization techniques, pulse measurements, signal processing algorithms.**

I. Introduction

X-RAY microcalorimeters respond electrically to the temperature change produced by the absorption of a photon. The electrical response is a brief drop in the bias current, a pulsed signal. Higher photon energies produce larger pulses. In the ideal case of a strictly linear sensor, pulses of any energy are perfectly scaled copies of each other, with the scale factor proportional to the photon energy. In the real world, pulse sizes and shapes can both depend on energy; "larger" pulses can mean pulses that have a larger peak amplitude but also last longer, if the sensor is driven past the small-signal limit.

The usual approach to the high-resolution analysis of microcalorimeter signals, however, assumes that all pulses have identical shapes. This assumption, along with others about the nature of the noise[1] leads to the procedure of *optimal filtering* [1]. One extracts from the signal data stream a record that contains N successive samples, with the pulse starting at some predetermined position in the record. The pulse size is then estimated by taking an inner product of the data record and the filter. The filter is thus a length-N weighting vector, one that is statistically optimal in a sense we shall see shortly.

We consider the problem of how to generalize the very successful framework of optimal filtering to improve energy resolution in cases where sensor nonlinearity is a serious obstacle. We show how to assess the expected resolution in any given data set under the new and standard analysis methods. This work considers only pulses isolated in time, but we fully expect that a deeper understanding of how to handle sensor nonlinearity will also be critical to the analysis of piled-up pulses. In short, we ask how to know when optimal filtering is, in fact, *not* optimal, and what can be done about it.

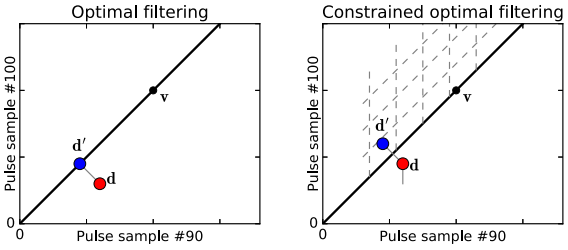


Fig. 1. A geometric view of optimal filtering without (*left*, Eq. (1)) and with constraints (*right*, Eq. (2)). Constrained filtering implies projection into a low-dimensional subspace (dashed coordinate system) whose coordinates are, in general, oblique. The coordinate giving distance parallel to the solid line is the pulse height, h; other coordinates correspond to nuisance terms. Only two arbitrary samples out of N are plotted, but one should imagine the points and lines to exist in an N-dimensional space.

II. Optimal Filtering

First, we review optimal filtering under the usual assumption of fixed pulse shape. We will use the term *pulse height* to describe any estimate of the "size" of a pulse, broadly construed, and not literally the peak value of the signal. The absolute calibration of pulse heights into energy is a separate procedure, beyond the scope of this paper.

To construct the filter $\boldsymbol{f}$ that estimates pulse height, we need estimates of both the signal shape $\boldsymbol{v}$ (of length N) and the noise covariance matrix[2] $\mathbf{R}$ of size $N \times N$. Then we define the filter and the pulse height estimator, h, to be

$$\boldsymbol{f}^{\mathrm{T}} \equiv [\boldsymbol{v}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{v}\,]^{-1}\boldsymbol{v}^{\mathrm{T}}\mathbf{R}^{-1} \text{ and } h \equiv \boldsymbol{f}^{\mathrm{T}}\boldsymbol{d}. \qquad (1)$$

Geometrically, we can interpret h as shown in Fig. 1 if we assume that the noise is white (i.e., $\mathbf{R} = \sigma^2\mathbf{I}$): the filter projects $\boldsymbol{d}$ onto the line passing through $\boldsymbol{v}$ and the origin, and h gives the coordinate along that line of the projected point.

When the noise is not white, then the projection is not the usual orthogonal projection. The $\mathbf{R}^{-1}$ factors ensure that the projection minimizes the *Mahalanobis distance*, rather than the Euclidean distance. Mahalanobis distance between a point and the center of some Gaussian distribution answers "how

Manuscript received September 6, 2016; accepted November 21, 2016. Date of publication December 8, 2016; date of current version January 11, 2017. This work was supported in part by the U.S. Department of Energy's Office of Basic Energy Sciences, in part by the NIST Innovations in Measurement Science proagram, and in part by NASA.

The authors are with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: joe.fowler@nist.gov; alpert@boulder.nist.gov; doriese@boulder.nist.gov; james.hays-wehle@nist.gov; youngil.joe@nist.gov; kelsey.morgan@nist.gov; galen.oneil@nist.gov; carl.reintsema@nist.gov; dan.schmidt@nist.gov; joel.ullom@nist.gov; daniel.swetz@nist.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASC.2016.2637359

[1]To wit, we assume that the noise is stationary, follows a multivariate Gaussian distribution of known covariance, and is independent of the signal.

[2]I.e., R_{ij} is the expected noise covariance between samples i and j.

Alpert, Bradley; Doriese, William; Fowler, Joseph; Hays-Wehle, James; Joe, Young Il; Morgan, Kelsey; O'Neil, Galen; Schmidt, Daniel; Swetz, Daniel; Ullom, Joel.
"When 'Optimal Filtering' Isn't."
Paper presented at Applied Superconductivity, Denver, CO, United States. September 5, 2016 - September 9, 2016.

many standard deviations away?" One way to view the minimization of Mahalanobis distance is that it reduces to Euclidean distance if both data $\boldsymbol{d}$ and pulse model $\boldsymbol{v}$ are first "whitened". That is, we replace $\boldsymbol{v} \rightarrow \mathbf{W}\boldsymbol{v}$ and $\boldsymbol{d} \rightarrow \mathbf{W}\boldsymbol{d}$, where $\mathbf{W}$ is any matrix that satisfies $\mathbf{R}^{-1} = \mathbf{W}^{\mathrm{T}}\mathbf{W}$. If it is understood that the projections take place in this "noise-whitened space," then we can maintain the geometric view of the filtering operations. The filter defined here is the optimal linear estimator of pulse size, in that it is both the minimum-variance unbiased linear estimator and also the maximum-likelihood estimator, but only under the assumption of strictly constant pulse shapes.

A. *Constrained Optimal Filters*

It is often necessary to allow for a model in which pulse records are linear combinations of a few components, the usual pulse shape *plus* one or more additional "nuisance" components. We often find the best results in allowing for two additional components: a constant offset and a term $\mathrm{d}\boldsymbol{v}/\mathrm{d}t$, which represents a first-order correction to the pulse shape $\boldsymbol{v}$ for small variations in the pulse arrival time t. We can write the model's n linear components as the columns of the matrix $\mathbf{M}$; our goal is to find the vector $\boldsymbol{p}$ such that $\mathbf{M}\boldsymbol{p}$ estimates the observed signal $\boldsymbol{d}$. In this case, the optimal filter becomes [2]

$$\mathbf{F}^{\mathrm{T}} \equiv [\mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}]^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}, \text{ and } \boldsymbol{p} = \mathbf{F}^{\mathrm{T}}\boldsymbol{d}. \qquad (2)$$

One element of $\boldsymbol{p}$ is h, the estimated pulse height of signal $\boldsymbol{d}$. If only the pulse height is needed, then we can use a unit vector $\hat{\boldsymbol{e}}_1$ to select the appropriate column of $\mathbf{F}$ and discard the nuisance components. We call this column the *constrained optimal filter* $\boldsymbol{f}_c = \mathbf{F}\hat{\boldsymbol{e}}_1$, where pulse height is estimated as before: $h_c = \boldsymbol{f}_c^{\mathrm{T}}\boldsymbol{d}$.

We can interpret $\boldsymbol{d}' \equiv \mathbf{M}\mathbf{F}^{\mathrm{T}}\boldsymbol{d}$ as a projection of the point $\boldsymbol{d}$ onto the n-dimensional subspace spanned by the columns of $\mathbf{M}$. Again, this projection will be the usual Euclidean projection only if the noise is white. For a more general noise matrix $\mathbf{R}$, we have instead that $\boldsymbol{d}'$ is the point in the subspace with the smallest Mahalanobis distance to the measured point $\boldsymbol{d}$. In general, the matrix $\mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}$ is not diagonal, and therefore $\boldsymbol{p}$ provides *oblique* coordinates in the subspace. Oblique coordinates imply higher and correlated uncertainties.

The constrained filter pulse height h_c is the optimal linear estimator of pulse size when the contribution of the $n-1$ unknown nuisance components must also be allowed to vary. It is optimal in that it is both the minimum-variance unbiased linear estimator and also the maximum-likelihood estimator, but again only if pulse shapes are strictly constant [3].

Assuming a known data noise covariance matrix $\mathbf{R}$, we can predict the energy resolution as a function of pulse energy. The energy uncertainty δE is a direct result of the pulse height uncertainty (let $\mathbb{E}\cdot$ represent expected values):

$$(\delta h)^2 = \mathbb{E}[h^2] - \mathbb{E}[h]^2 = \boldsymbol{f}_c^{\mathrm{T}}\mathbf{R}\boldsymbol{f}_c; \qquad (3)$$

$$\delta E \approx \delta h \,/\, \mathbb{E}[\mathrm{d}h/\mathrm{d}E]. \qquad (4)$$

To estimate $\mathrm{d}h/\mathrm{d}E$, we need to generalize the model pulse and allow $\boldsymbol{v} = \boldsymbol{v}(E)$ to be a function of energy. Then

$$\mathbb{E}[h(E)] = \boldsymbol{f}_c^{\mathrm{T}}\boldsymbol{v}(E), \qquad (5)$$

$$\mathbb{E}[\mathrm{d}h/\mathrm{d}E] = \boldsymbol{f}_c^{\mathrm{T}}\,\mathrm{d}\boldsymbol{v}/\mathrm{d}E, \qquad (6)$$

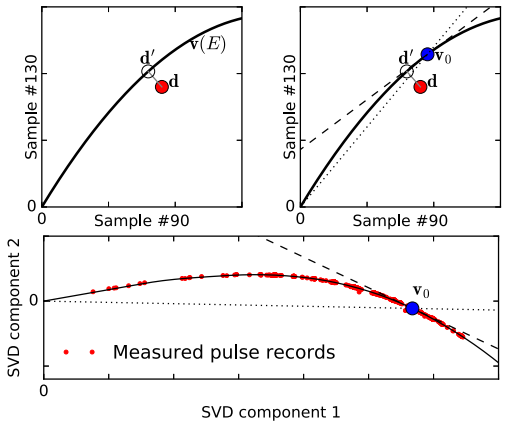


Fig. 2. When sensors are nonlinear, their response $\boldsymbol{v}(E)$ can be a curve (*top left*). Here, two arbitrary samples out of N are shown. To find the point, $\boldsymbol{d}'$, on the curve nearest to any given measurement, $\boldsymbol{d}$, is not the strictly linear operation of Fig. 1. *Top right:* given a reference pulse $\boldsymbol{v}_0$, one can approximate the nearest point to $\boldsymbol{d}$ by linear operations, by projecting $\boldsymbol{d}$ onto either the secant line (dotted) or the tangent line passing through $\boldsymbol{v}_0$ (dashed). These steps correspond, respectively, to optimal filtering and to tangent filtering. *Bottom:* Measurements up to 9 keV, described in Section V. The coefficients of the two leading singular vectors are shown for some 1000 pulses (dots), as well as a smooth model (solid curve) and the lines onto which data are projected by optimal filtering (dotted) and tangent filtering (dashed).

and so the expected energy resolution[3] is

$$\delta E = \sqrt{\boldsymbol{f}_c^{\mathrm{T}}\mathbf{R}\boldsymbol{f}_c} \,/\, \boldsymbol{f}_c^{\mathrm{T}}(\mathrm{d}\boldsymbol{v}/\mathrm{d}E). \qquad (7)$$

The numerator is a fixed quantity because the pulse height uncertainty δh is independent of pulse energy.[4] The energy dependence appears only in the denominator, in two ways. For one, the pulse size often grows less rapidly at high energies; also, the pulse shape can change with energy, varying the angle between the vectors $\boldsymbol{f}_c$ and $\mathrm{d}\boldsymbol{v}/\mathrm{d}E$.

III. Noise-Free Pulses Describe a Curve in $\mathbb{R}^N$

When pulse shapes change with energy, the assumptions that lead to the constrained optimal filter have failed. Will correcting these assumptions yield better energy resolution?

Consider the idealized, noise-free pulse shape as a function of energy $\boldsymbol{v}(E)$. This function describes a 1-dimensional curve in $\mathbb{R}^N$ (Fig. 2). Provided the noise is low, measured pulses will cluster in a cloud around this curve. To estimate the energy of a point $\boldsymbol{d}$ in the cloud, we should choose the "nearest" point on the curve (as before, we mean minimal Mahalanobis distance) and assign to $\boldsymbol{d}$ the energy E of that point on the curve.[5]

Many approaches to accommodate an energy-dependent pulse shape have previously been proposed. Fixsen *et al.* [4] have suggested the use of a discrete set of models at fixed energies; the model selected for a given pulse would be the model that produces the best fit or an appropriate linear interpolation between the two best models [5], [6]. Others have proposed use of the standard optimal filter to choose among models, followed

[3]This δE is a standard deviation; multiply by $\sqrt{8\ln 2}$ to get FWHM.

[4]This statement assumes that the noise does not depend on the current through the sensor, though there are physical reasons to expect some suppression of TES noise at the peak of a pulse.

[5]If the curve is parameterized by some uncalibrated quantity such as pulse height, then an additional calibration step is required, as with optimal filtering.

Alpert, Bradley; Doriese, William; Fowler, Joseph; Hays-Wehle, James; Joe, Young Il; Morgan, Kelsey; O'Neil, Galen; Schmidt, Daniel; Swetz, Daniel; Ullom, Joel.
"When 'Optimal Filtering' Isn't."
Paper presented at Applied Superconductivity, Denver, CO, United States. September 5, 2016 - September 9, 2016.

by interpolation [7]; a fully general treatment based on differential geometry [8]; and an expansion of the signal to linear order in small changes in energy [9]. A similar linearization of the response has also been considered for use with position-sensitive detectors [10].

IV. Tangent Filtering for Nonlinear Data

If the pulse shape $\boldsymbol{v}(E)$ is a slowly varying function of energy, then we can choose a reference point on the curve $\boldsymbol{v}_0 = \boldsymbol{v}(E_0)$ and linearize the curve in that vicinity. This will mean projecting data points $\boldsymbol{d}$ onto the line tangent to the curve at E_0. By contrast, optimal filtering in effect projects onto the secant line, the line between the origin and $\boldsymbol{v}_0$ (Fig. 2). The point projected onto the tangent line should be closer to the true $\boldsymbol{v}(E)$ curve than the same point projected onto the secant line. Whether this smaller residual yields improved energy resolution is a separate question, but it should offer clear advantages at least in the matter of identifying pathological pulse records (e.g., pulse pileup) via their rms residual.

We call this use of linearized signal *tangent filtering*, in that we use a filter that projects the data to the Mahalanobis-nearest point on the line tangent to $\boldsymbol{v}(E)$.

To effect tangent filtering, we operate on the data after subtracting from it the reference pulse $\boldsymbol{v}_0$. We must also replace the signal column in the n-column model matrix $\mathbf{M}$; instead of $\boldsymbol{v}_0$, the first column of $\mathbf{M}$ must be the tangent line $(\mathrm{d}\boldsymbol{v}/\mathrm{d}E)$. Call this new model matrix for the tangent-line model $\mathbf{M}_t$. Thus the tangent filter is

$$\mathbf{F}_t^{\mathrm{T}} \equiv [\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}_t\,]^{-1}\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}, \tag{8}$$

and it is applied to data $\boldsymbol{d}$ to estimate the contribution of the m components via

$$\boldsymbol{p} = \mathbf{F}_t^{\mathrm{T}}(\boldsymbol{d} - \boldsymbol{v}_0)$$

where the energy estimate is

$$\tilde{E} = \hat{\boldsymbol{e}}_1^{\mathrm{T}}\boldsymbol{p} + E_0.$$

Unlike optimal filtering, this result is already calibrated into energy (provided that the pulse curve $\boldsymbol{v}(E)$ is parameterized by a calibrated energy). This fact means that we can estimate energy resolution without needing to compute the $\mathrm{d}h/\mathrm{d}E$ conversion factor. The square of the rms resolution δE is:

$$\begin{aligned}(\delta E)^2 &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}\left\{\mathbb{E}[\boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}] - \mathbb{E}[\boldsymbol{p}]\mathbb{E}[\boldsymbol{p}]^{\mathrm{T}}\right\}\hat{\boldsymbol{e}}_1 \\ &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}\mathbf{F}_t^{\mathrm{T}}\left\{\mathbb{E}[\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}] - \mathbb{E}[\boldsymbol{d}]\mathbb{E}[\boldsymbol{d}]^{\mathrm{T}}\right\}\mathbf{F}_t\hat{\boldsymbol{e}}_1 \\ &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}\mathbf{F}_t^{\mathrm{T}}\mathbf{R}\mathbf{F}_t\hat{\boldsymbol{e}}_1 \\ &= \hat{\boldsymbol{e}}_1^{\mathbf{T}}[\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}_t]^{-1}\hat{\boldsymbol{e}}_1. \end{aligned} \tag{9}$$

The energy uncertainty is thus the square root of the $(1,1)$ component of the inverse of $\mathbf{A} \equiv [\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}_t]$. We have verified Eq. 9 and Eq. 7 by applying both types of filter to simulated noise records that have the appropriate covariance.

We can get a sense of this result in the case of $n = 2$ components. Let the columns of the noise-whitened model $\mathbf{W}\mathbf{M}_t$ have magnitudes a and b and form an angle θ. Then

$$\mathbf{A} = (\mathbf{M}_t^{\mathrm{T}}\mathbf{W}^{\mathrm{T}})(\mathbf{W}\mathbf{M}_t) = \begin{pmatrix} a^2 & ab\cos\theta \\ ab\cos\theta & b^2 \end{pmatrix},$$

$$\delta E = (a\sin\theta)^{-1} = (||\mathbf{W}\,\mathrm{d}\boldsymbol{v}/\mathrm{d}E||\,\sin\theta)^{-1}.$$

Thus, the energy resolution is the inverse of the norm of the noise-whitened and -scaled pulse derivative $(\mathrm{d}\boldsymbol{v}/\mathrm{d}E)$, times a geometric factor to penalize the model insofar as it has a non-orthogonal second component. If $\mathbf{M}$ has $n \geq 3$ columns, the result is similar: $\delta E = g/a$ where $g \geq 1$ is a geometric factor that involves only the angles between noise-whitened components of the signal model. The geometric factor still penalizes the oblique coordinates, and it reduces to 1 when the coordinates are all strictly orthogonal.

The expected resolution in the cases of standard, constrained optimal filtering (Eq. 7) and tangent filtering (Eq. 9) can be compared for any given data set, so long as the following are known: the pulse size and shape as a function of energy $\boldsymbol{v}(E)$; the additional linear, nuisance components to be included in the model; and the noise covariance matrix $\mathbf{R}$. We find the resolution ratio to be a useful guide to when we can expect tangent filtering to yield an improved energy resolution.

The tangent filter certainly does not solve all problems that arise from sensor nonlinearity. Consider the special case that pulse sizes depend nonlinearly on energy, but the pulse shape is constant: $\boldsymbol{v}(E) = s(E)\boldsymbol{v}_0$. In this case, the corresponding columns of $\mathbf{M}$ and $\mathbf{M}_t$ are parallel, and both Eq. 7 and 9 reduce to the same result. Importantly, the result scales as

$$\delta E \propto (\mathrm{d}s/\mathrm{d}E)^{-1}.$$

The usual form of pulse nonlinearity is a compression, a reduction in $\mathrm{d}s/\mathrm{d}E$ as E grows; we see that this causes energy uncertainty to grow with energy. While tangent filtering is capable of recovering energy resolution lost to changes in pulse *shape*, it cannot help with resolution lost because of pulse size compression.

V. Case Study: Kα Emission up to 8 keV

Here we present a case study in which tangent filtering produces the predicted effect, a 15 % improvement in energy resolution. The data are one TES's measurements of the K-line emission of several 3d transition metals, with energies in the range of 4 keV to 8 keV. We assess energy resolution by fitting for the width of the Gaussian energy-response function in measurements of two-peaked K$\alpha_{1,2}$ line complexes.

A. Pulse Modeling

Tangent filtering requires a model for the pulse size and shape versus energy, $\boldsymbol{v}(E)$. It should accurately capture $\mathrm{d}\boldsymbol{v}/\mathrm{d}E$, at least near the energies of interest. In contrast, the usual optimal filter analysis does not require such a model—one needs only a single pulse model (typically, the average over many measured pulses at a range of energies, which might well not match the pulse shape for any single energy). Therefore, we think it valuable to explain how we have constructed this model.

So that we can assign a tentative energy label to each pulse, we compute the *pulse rms* for each (that is, the root mean square of the measured samples, after a pre-pulse mean value is subtracted). Using the observed spectrum of pulse rms values, we find an appropriate calibration curve that converts the pulse rms into a preliminary estimate of energy.

To model $\boldsymbol{v}$ is a challenge. We need to estimate its energy derivative at each energy of interest, while the data available tend to cluster near a limited set of energies (here, the Kα and Kβ line energies). We start with the observation that the curve

Alpert, Bradley; Doriese, William; Fowler, Joseph; Hays-Wehle, James; Joe, Young Il; Morgan, Kelsey; O'Neil, Galen; Schmidt, Daniel; Swetz, Daniel; Ullom, Joel.
"When 'Optimal Filtering' Isn't."
Paper presented at Applied Superconductivity, Denver, CO, United States. September 5, 2016 - September 9, 2016.

TABLE I
ENERGY RESOLUTIONS (FWHM, IN eV), PREDICTED AND MEASURED

Emission Line	Counts in peak	Predicted resolution Optimal	Predicted resolution Tangent	Observed resolution Optimal	Observed resolution Tangent
Mn Kα	4500	3.38	3.19	4.58	4.14
Fe Kα	3200	3.56	3.29	4.76	4.36
Co Kα	2500	3.78	3.42	4.53	4.17
Ni Kα	6700	4.10	3.56	4.72	3.99
Cu Kα	7800	4.36	3.69	4.92	4.20

$\boldsymbol{v}(E)$ is in practice approximately confined to a low-dimensional subspace of the full $\mathbb{R}^N$. If we can identify a basis to span this subspace, of dimension $N_{\rm sub} \ll N$, then we need not express $\boldsymbol{v}(E)$ as N separate functions of E; a mere $N_{\rm sub}$ coefficients of that basis can approximate it instead.

Many choices of basis are possible. In the present work, we use the nine leading right singular vectors from the singular-value decomposition (SVD) of a matrix whose columns are many (10^4) pulse records. We model the nine coefficients versus energy via least-squares-approximating cubic splines with ten knots between 2 keV and 9 keV, the knots more concentrated where the data are most dense (in the range of 5 keV to 7 keV). Fig. 2 shows data records and the smooth model projected into the two dimensions with the highest singular values.

B. Filtering Results

We apply Equations 2 and 8 to compute the constrained optimal filter $\mathbf{F}$ and the tangent filter $\mathbf{F}_t$. In the latter case, we compute three tangent filters, appropriate to the energies of the Kα lines of Mn, Co, and Cu. To reduce systematic dependence on sub-sample arrival time, we smooth the four filters with a one-pole low-pass filter ($f_{3\,\rm dB} = 12$ kHz), after which we apply each filter to each pulse. The low-pass filtering and a gain adjustment to correct for cryogenic temperature drifts are standard steps in our best practices for achieving high energy resolution from TES microcalorimeters [11]. Finally, this TES is susceptible to cross-talk from others in the array. We cut any pulses that coincide with pulses seen in the array's other sensors (some 50% of them), improving either method's energy resolution by nearly 1 eV.

The cut, filtered, and corrected data are combined into energy spectra. We use maximum-likelihood fits [12] to the known Kα line shapes [13] to extract the estimated energy resolution (Table I). The statistical uncertainty in the resolution fits is typically ± 0.2 eV for the Mn, Ni, and Cu fits, or ± 0.4 eV at the Fe and Co lines, which contain fewer photons. Besides the Gaussian resolution shown in the table, these detectors exhibit an exponential tail to low energies, a result of long-lived thermal states in the bismuth x-ray absorber. The resolutions, predicted or measured, increase with energy due to compression.

The observed resolution columns show that tangent filtering improves the resolution relative to the standard optimal filtering by 0.7 eV at the higher energies, and by at least 0.4 eV at all energies studied. This is a successful application of the tangent filtering method to actual TES measurements in the presence of a pulse shape that changes with energy. We attribute the remaining predicted-observed discrepancy to undetected cross-talk—some active sensors were not being recorded in this measurement—and to arrival-time systematics that arise because of the fast TES rise time (35 μs).

VI. FUTURE OUTLOOK

We have described a process to model pulse shapes and enable tangent filtering, as well as one measurement in which it provided substantially improved energy resolution: from 4.9 eV to 4.2 eV at the Cu Kα line.

Our experience suggests that improved energy resolution is not universally available, unfortunately. The angle between vectors $\boldsymbol{v}$ and $\mathrm{d}\boldsymbol{v}/\mathrm{d}E$ is generally quite small unless the TESs are operated near to reaching their normal-state resistance. We considered a variety of other data for this test; finding that the predicted resolutions improved by only a few percent, we did not attempt tangent filtering.

One question we have not explored here is one of global calibration, beyond the narrow range of a Kα complex. If we have multiple filters based on the tangent to $\boldsymbol{v}(E)$ at various E, then each will yield a different estimate of pulse energy; a procedure is needed to select or interpolate among them if an energy spectrum is to be measured over a wide range.

We plan further work to help us understand when tangent filtering improves the energy resolution, and to streamline the modeling process for routine work on large data sets produced by arrays of hundreds of detectors and over a variety of energy spectra. If such modeling succeeds, then in cases where "optimal filtering" is not actually optimal, the tangent filtering method presented here should provide a valuable improvement in the energy resolution achieved from microcalorimeter sensors.

REFERENCES

[1] A. E. Szymkowiak, R. L. Kelley, S. H. Moseley, and C. K. Stahle, "Signal processing for microcalorimeters," *J. Low Temp. Phys.*, vol. 93, no. 3/4, pp. 281–285, Nov. 1993.

[2] B. K. Alpert *et al.*, "Note: Operation of gamma-ray microcalorimeters at elevated count rates using filters with constraints," *Rev. Sci. Instrum.*, vol. 84, no. 5, May 2013, Art. no. 6107.

[3] J. W. Fowler *et al.*, "Microcalorimeter spectroscopy at high pulse rates: A multi-pulse fitting technique," *Astrophys. J. Suppl. Ser.*, vol. 219, no. 2, Aug. 2015, Art. no. 35.

[4] D. J. Fixsen, S. H. Moseley, B. Cabrera, and E. Figueroa-Feliciano, "Optimal fitting of non-linear detector pulses with nonstationary noise," in *Proc. 9th Int. Workshop Low Temp. Detectors*, vol. 605, no. 1, pp. 339–342, Feb. 2002.

[5] D. Fixsen, S. Moseley, and B. Cabrera, "Pulse estimation in nonlinear detectors with nonstationary noise," *Nuclear Instrum. Methods Phys. Res. Sec. A*, vol. 520, pp. 555–558, 2004.

[6] B. Shank *et al.*, "Nonlinear optimal filter technique for analyzing energy depositions in TES sensors driven into saturation," *AIP Adv.*, vol. 4, no. 1, Nov. 2014, Art. no. 117106.

[7] C. Whitford, W. B. Tiest, and A. Holland, "Practical considerations in optimal filtering of TES signals," *Nuclear Instrum. Methods Phys. Res. Sec. A, Accelerators, Spectrometers, Detectors Assoc. Equipment*, vol. 520, no. 1, pp. 592–594, 2004.

[8] D. J. Fixsen, S. H. Moseley, T. Gerrits, A. E. Lita, and S. W. Nam, "Optimal energy measurement in nonlinear systems: An application of differential geometry," *J. Low Temp. Phys.*, vol. 176, no. 1/2, pp. 16–26, Mar. 2014.

[9] P. Peille *et al.*, "Performance assessment of different pulse reconstruction algorithms for the ATHENA X-ray integral field unit," *SPIE Astron. Telesc. Instrum.*, vol. 9905, pp. 99 055W–99 055W–14, Jul. 2016.

[10] S. J. Smith, "Implementation of complex signal-processing algorithms for position-sensitive microcalorimeters," *Nuclear Instrum. Methods Phys. Res. Sec. A*, vol. 602, no. 2, pp. 537–544, Apr. 2009.

[11] J. W. Fowler *et al.*, "The practice of pulse processing," *J. Low Temp. Phys.*, vol. 184, no. 1, pp. 374–381, Jul. 2016.

[12] J. W. Fowler, "Maximum-likelihood fits to histograms for improved parameter estimation," *J. Low Temp. Phys.*, vol. 176, no. 3, pp. 414–420, Aug. 2014.

[13] G. Hölzer, M. Fritsch, M. Deutsch, and J. Härtwig, "K$\alpha_{1,2}$ and K$\beta_{1,3}$ x-ray emission lines of the 3d transition metals," *Phys. Rev. A*, vol. 56, no. 6, pp. 4554–4568, 1997.

Alpert, Bradley; Doriese, William; Fowler, Joseph; Hays-Wehle, James; Joe, Young Il; Morgan, Kelsey; O'Neil, Galen; Schmidt, Daniel; Swetz, Daniel; Ullom, Joel.
"When 'Optimal Filtering' Isn't."
Paper presented at Applied Superconductivity, Denver, CO, United States. September 5, 2016 - September 9, 2016.

Refractive index measurements of Ge

John H. Burnett*[a], Simon G. Kaplan[a], Erik Stover[b], and Adam Phenis[c]

[a]National Institute of Standards and Technology, Physical Measurement Laboratory, 100 Bureau Dr., Gaithersburg, MD USA 20899; [b]M3 Measurement Solutions, Inc., 938 S. Andreasen Dr., Suite I, Escondido, CA USA 92029; [c]AMP Optics, 13308 Midland Rd., #1304, Poway, CA USA 92074

ABSTRACT

A program has been started at NIST to make high-accuracy measurements of the infrared (IR) index properties of technologically important IR materials, in order to provide the IR optics community with updated values for the highest-quality materials now available. For this purpose, we designed and built a minimum-deviation-angle refractometry system enabling diffraction-limited index measurements for wavelengths from 0.12 μm to 14 μm. We discuss the apparatus and procedures that we use for IR measurements. First results are presented for germanium for the wavelength range from 2 μm to 14 μm, with standard uncertainties ranging from 2×10^{-5} near 2 μm to 8×10^{-5} near 14 μm. This is an improvement by about an order of magnitude of the uncertainty level for index data of germanium generally used for optic design. A Sellmeier formula fitting our data for this range is provided. An analysis of the uncertainty is presented in detail. These measurements are compared to previous measurements of Ge.

Keywords: index, index of refraction, refractive index, dispersion, refractometry, germanium, Ge, infrared

1. INTRODUCTION

Since germanium (Ge) first began to be developed for transistor and infrared (IR) optics applications in the late 1940s, the material quality has continually improved to meet increasingly stringent demands. For optics applications, this has driven the need for more accurate optical properties measurements, especially the index of refraction in the high IR transmission wavelength region (2 μm to 14 μm for dry air), to meet the requirements of optical design. Ge measurements by various methods in the 1950s reported absolute index measurement accuracies ranging from high 10^{-3} to a few parts in 10^{-4}, with results between labs differing by as much as 1 percent.[1,2,3,4] These materials had various doping and defect levels. The measurements by Salzberg and Villa[3,4] in 1957 and 1958 for the range 2 μm to 16 μm have been used extensively for IR optics design for two decades since their publication, and to some extent even today.

By the mid-1970s, driven by the importance of Ge for IR applications, material quality and index measurement methods had improved to the point where reported index measurements for the range of 2 μm to 14 μm generally had stated uncertainty estimates in the mid to low 10^{-4} range.[5,6] In particular, Icenogle *et al.*[5] reported measurements for the refractive indexes of Ge over the range 2.5 μm to 12 μm and their variations with temperature over the ranges 95 K to 298 K, with conservative estimates of the their standard uncertainties of $\pm 6 \times 10^{-4}$. These differed from previous measurements, e.g., those of Salzberg and Villa[4], by as much as 5×10^{-3}, when account is taken of the different measurement wavelengths and temperatures used. Icenogle *et al.*[5] suggested that the difference may be due to different sample impurities. Edwin *et al.*[6] reported measurements of a Ge sample at two laboratories over a subrange 8 μm to 14 μm by different deviation-angle methods, with a discrepancy of 3×10^{-4} between labs. These results differed from those of Icenogle *et al.* by $\leq 10 \times 10^{-4}$ over the range of overlap (with the exception of longest wavelength index). In 1980, Li[7] reviewed and tabulated all the measured index data on Ge and Si since 1949. Based on an analysis of all the data, he recommended index values for Ge for wavelength and temperature in the ranges 1.0 μm to 18 μm and 100 K to 550 K, in the form of an expression with the wavelength dispersion given by a single $1/\lambda^2$ term. The stated standard uncertainty for the index was $\pm 2 \times 10^{-3}$ over the wavelength and temperature range, consistent with the uncertainties claimed for the 1970s measurements discussed above.

*john.burnett@nist.gov; phone 301-975-2679

Burnett, John; Kaplan, Simon.
"Refractive index measurements of Ge."
Paper presented at SPIE Optics and Photonics, San Diego, CA, United States. August 28, 2016 - September 1, 2016.

Since the late 1970s, the results of Icenogle *et al.*[5], and to some extent the previous results of Salzberg and Villa[3,4], have been widely used in IR optics design, including incorporation in commercial optics design software. These values have been recommended in well-established index data reference books, such as the Handbook of Optical Constants of Solids[8] and the Handbook of Infrared Optical Materials.[9] With no substantial additions to these index data over the following 30 years, they remain the values IR optical designers must rely on, accepting a measurement uncertainty for each sample of at best (1 to 2) $\times 10^{-3}$ and sample-to-sample variation up to an order of magnitude higher.

There have been a number of different methods used to measure the index of bulk materials, including techniques based on refraction angle, interference in slabs, surface reflection, and ellipsometry. For the transparent region of IR materials, the methods reporting the highest accuracy are variations of the minimum-deviation-angle refractometry method, based on measuring the angle of refraction through a prism of the material set at the angle of minimum deviation, discussed below. The accuracy of this approached is limited by the diffraction of the beam through the prism. For a prism about 3 cm on a side, it is straightforward to calculate that the diffraction-limited index uncertainty is approximately 5×10^{-7} in the ultraviolet, which has been verified.[10] This diffraction-limited uncertainty should scale with wavelength approximately two orders of magnitude to the mid 10^{-5} range in the mid-IR. This is an order of magnitude smaller than the uncertainties reported for Ge discussed above. Thus, with an optimized refractometer design, an order of magnitude improvement in index accuracies should be achievable. Besides the inherent value of this higher accuracy for more precise IR optical design, this level of accuracy would be of value to sort out index variations in samples of commercial material, feeding back to material fabricators to improve material quality.

Over the last several years a program has been established at NIST to pursue these goals for materials of importance to the IR optics community. In this paper we discuss the minimum-deviation-angle refractometry system developed at NIST and the procedures designed to achieve diffraction-limited index measurements in the IR out to 14 μm. Achieving this performance requires substantially reducing numerous sources of uncertainty compared to previous measurements, in some cases by an order of magnitude or more, as will be discussed. We will present our index measurements of a sample of high-quality Ge for wavelengths from 2 μm to 14 μm at the temperature of 22.000 °C. We fit these data to a Sellmeier formula and compare the values to previous measurements.

2. MEASUREMENT

The well-established minimum-deviation-angle refractometry method is based on determining the angle $\delta(\lambda)$ that a collimated beam of known wavelength λ refracts through a prism of known apex angle α.[11,12] If the prism is oriented so that the beam travels symmetrically through the prism with respect to the apex angle, then the refraction angle (deviation angle) is at a minimum and the refractive index of the prism material $n_{mat}(\lambda)$ relative to the index of the gas surrounding the prism $n_{gas}(\lambda)$ is given by the following formula:

$$n_{rel}(\lambda) = n_{mat}(\lambda)/n_{gas}(\lambda) = \sin\left[\left(\alpha + \delta(\lambda)\right)/2\right]/\sin(\alpha/2). \quad (1)$$

A substantial benefit of measuring the deviation angle $\delta(\lambda)$ at the angle of minimum deviation is that the measurement is at an extremum condition, making it fairly insensitive to the alignment of the prism, a substantial source of error. Due to the conceptual simplicity of the index determination, not depending on any models other than Snell's law, it is relatively secure from subtle sources of systematic errors. The accuracy of the determination of $n_{mat}(\lambda)$ simply depends on the accuracy of the of the determination of the parameters, λ, $n_{gas}(\lambda)$, α, and $\delta(\lambda)$. The uncertainties in α and $\delta(\lambda)$ depend in part on the geometric and surface properties of the prism, along with the prism index homogeneity and isotropy. We give some discussion on those aspects of our implementation of the minimum deviation angle technique relevant to the uncertainties of these various components.

2.1 NIST Minimum-Deviation-Angle Refractometer

A custom minimum deviation angle refractometer system used for high-accuracy index measurements in the visible and ultraviolet, was modified with sources, detectors, and-all reflective optics appropriate for operation in the IR. A schematic diagram of the system is shown in Figure 1. The narrow-band IR light for the $\delta(\lambda)$ measurements was generated by filtering radiation from a broadband 1200 °C blackbody source by a grating monochromator and from the a 3.39 μm HeNe

laser. The grating orders were separated by the prism sample. The entire path of the radiation from the source to the detector was enclosed in sealed housings purged with dry N_2 gas, to enable measurements in the low transmission regions of air with substantial H_2O and CO_2 content.

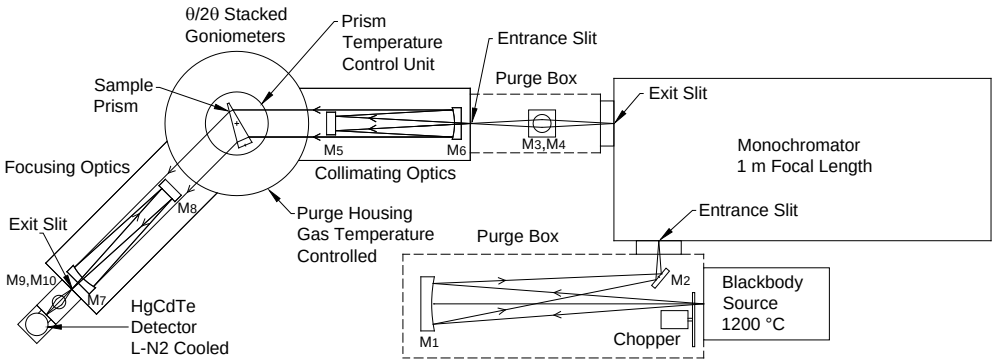


Figure 1. Schematic diagram of the top view of the NIST minimum-deviation-angle refractometer system with all reflective optics. Mirror M1 collects radiation from the blackbody source and images it, via M2, into a monochromator. The output of the monochromator is imaged into the higher refractometer entrance slit via a pair of parabolic mirrors M3 and M4. The radiation through the entrance slit is turned and angled down by plane mirror M5 and collimated by spherical mirror M6. The collimated beam is refracted by the prism and imaged to the exit slit by spherical mirror M7 and plane mirror M8. This radiation is re-imaged into the HgCdTe detector by a pair of parabolic mirrors M9 and M10.

The radiation from the blackbody source was chopped and imaged into a McPherson 2051 monochromator with a focal length of 1 m. A grating with 100 grooves per millimeter, blazed near 9 μm was used. Appropriate grating orders were chosen for each wavelength desired to ensure the grating was used near the blaze angle for high efficiency. The wavelength output of the monochromator was calibrated with the first six orders of HeNe 633 nm laser and the first four orders of a 3.39 μm HeNe laser. The small wavelength non-linearity of the monochromator was removed by the generated calibration curve. The wavelength was accurate to within ±0.2 nm for the full 2 μm to 14 μm range. The monochromator slit widths were chosen for each wavelength to be as large as possible to maximize the signal relative to the background, while ensuring that the bandwidth contributed less than 10% of the overall uncertainty budget for the index.

The narrow-band radiation from the monochromator was collected and focused into the entrance slit of the refractometer with a focal length of 0.5 m, by a pair of parabolic mirrors. The entrance and exit slit widths of the refractometer were chosen as above to maximize signal without degrading the index uncertainty above the diffraction limit. A folding mirror and spherical mirror collimated the radiation to the prism, and the refracted radiation was focused onto the exit slit by spherical mirror and folding mirror. A ray trace analysis of system showed that aberrations from using spherical mirrors for the slit height of 4 mm, was less than the diffraction spread. The optics of the refractometer had a f-number of 10.

The prism sample holder was mounted on a pair of stacked computer-controlled goniometers. The bottom goniometer, which holds the detector arm, had an encoder with absolute accuracy of ±0.2 arcseconds. This limits the accuracy of the deviation angle $\delta(\lambda)$. The top goniometer was controlled by the drive program to ensure that after initial minimum deviation alignment, the minimum deviation condition of the prism was maintained. The prism tilts were aligned by an autocollimator, and the minimum deviation alignment was determined by multiple scans at various prism angles at 3.39 μm.

Accurate temperature control for the prism at 22.000 °C was achieved by a feedback control loop in the following fashion. The prism was mounted between two hollow copper plates. Two water baths with water flow to the copper plates were maintained to 0.01 °C at temperatures less than one °C above and below the set point temperature of 22.000 °C. Water from the two baths was mixed a few inches before entering the copper plates. The mixing of the two fluids was

accomplished by a mixing value, with the mixing ratio determined by the error in the sample temperature from the set point, in a PID control loop. The sample temperature was determined by four platinum resistance thermometers mounted on the copper plates. These were calibrated by a standards-grade platinum resistance thermometer calibrated against the triple point of water and the melting point of gallium, according to the ITS-90 international temperature standards protocol.[13] The standard uncertainty of the calibration of each of the thermometers used was 0.002 °C. During temperature control operation, all four thermometers were monitored and they were always within 0.002 °C of each other. The temperature control loop kept all four thermometers oscillating within the range 22.000 ±0.005 °C. The temperature of the nitrogen gas was controlled by a second, similar PID control loop. The gas temperature within the sample housing was monitored by two calibrated platinum resistance thermometers and was controlled within the range 22.000 ±0.020 °C.

The radiation exiting the refractometer was detected by a liquid-nitrogen-cooled HgCdTe detector using phase sensitive (lock-in) detection. The deviation angle for each wavelength $\delta(\lambda)$ was determined by introducing each wavelength in turn into the entrance slit of the refractometer and scanning the detector arm, with the prism angle scanning at half the angular rate to ensure that the minimum-deviation condition was maintained. For each wavelength, the encoder-determined angle of the center of the peak (obtained by fitting) relative to that of the un-deviated beam, determined the deviation angle $\delta(\lambda)$. Measurements were made on both sides of prism repeatedly to cancel effects of absolute angle drifts. The uncertainty of the center of each peak was estimated conservatively to be 10% of the FWHM (full-width at half-peak-maximum). This uncertainty dominated the total estimated uncertainty for all wavelengths. The peak widths broadened as the wavelength increased, as expected from diffraction. Thus the total uncertainty increased as the wavelength increased, as discussed in Section 3.1.

2.2 Ge Sample

The Ge material measured was made into a prism with sides 39 mm long and 29 mm high, and an apex angle of approximately 15 degrees. Both prism transmitting surfaces had figure errors less than 5 nm RMS (root mean squared) (λ/120 RMS at 633 nm) over the 90% clear aperture. The apex angle was determined by use of stacked indexing tables and an autocollimator, making angle difference measurements on both sides of the prism with all combinations of index angles separated by 15 degree intervals. The standard deviation of all the measurements from the various indexing table angle combinations was 0.12 arcseconds. Including other sources of error, such as the surface figure errors, gave a standard uncertainty of the apex angle of 0.2 arcseconds.

The sample resistivity, determined by a 4-probe measurement, was 17 Ohm-cm, within the range about 10 Ohm-cm to 40 Ohm-cm generally targeted for optimal Ge transmission properties. The index homogeneity was measured on a 10 mm thick, 27 mm diameter Ge disk, taken from a location in the Ge ingot close to where the prism material was taken. The measurements were made at 3.39 μm, using a M3 Measurements Solutions MWAVE 339 IR phase-shifting interferometer, with a calibrated accuracy of 0.002 λ RMS wavefront error at 3.39 μm.[14] Due to the very high temperature dependence of the index of 3.96×10^{-4} /°C,[5] there was concern about temperature gradients in the sample giving rise to index inhomogeneities during the characterization. For the measurement, the sample was mounted between two large, copper blocks with thermally-connecting side plates and apertures. The result was 9.1×10^{-6} RMS, 72.8×10^{-6} PV (peak to valley). Due to possible residual temperature gradient effects, this represents an upper bound for the grown-in inhomogeneity.

We measured the transmittance through the center of the face of the prism with a 6 mm diameter aperture, using a Fourier-transform spectrometer and an integrating sphere to collect the transmitted light.[15] We found that there was no change in transmittance structure using smaller apertures. The results are shown in Figure 2. Spectra show a transmittance of about 44% through the range just above 2 μm through 11 μm. A calculation using our measured index values predicts a value of within 2% of this, the specific value depending on assumptions on what happens to the multiple reflections in this highly refracting prism window. This spectrum is similar qualitatively and quantitatively to transmittance spectra of high-quality Ge in the literature.[9] The sharp drop in transmittance to zero near 2 μm is due to the onset of electronic transitions at the semiconductor conduction band edge, and is the lower limit wavelength for use of Ge as thick transmitting windows and optics. To avoid the distorting effects of high absorption on the index measurements below 2 μm, 2.25 μm was used as our lowest measurement wavelength. The absorption structure at 12 μm and longer wavelengths is due to phonon absorption. The upper limit of the measurements was 14 μm.

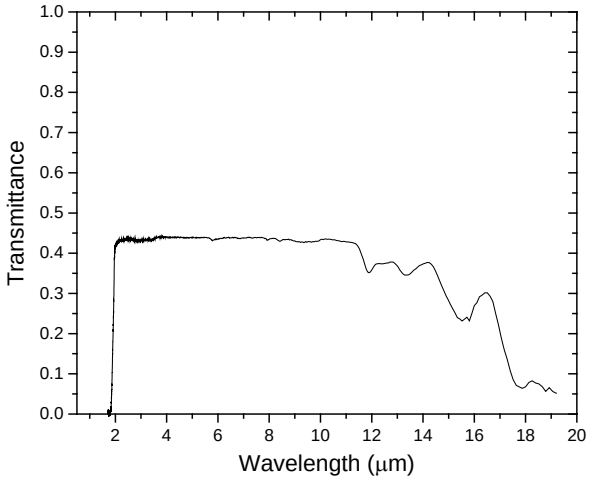


Figure 2. Transmittance versus wavelength of the 15 degree Ge prism measured using a Fourier-transform spectrometer and an integrating sphere. The illumination was apertured to 6 mm diameter at the center of the prism, with the first face at 23 degrees to the illumination and the back face 8 degrees to the illumination.

2.3 Measurement Procedures

Deviation angle measurements were made at minimum deviation, transmitting through the prism from both directions by rotating the prism around and measuring the deviations on both sides of the undeviated beam. This required establishing minimum deviation settings on each side by multiple index scans as a function of prism position. On each side, deviation angle measurements were made at 17 wavelengths shown in Table 1, within the range 2.25 μm to 14 μm. The deviation angle at each wavelength was determined by the average of the deviation angle from both sides. This was repeated 5 times. During the measurements, the temperature of the prism and the N_2 gas in the sample purge chamber were continuously monitored, but because they were controlled to 22.000 C, there was no need to make temperature corrections. The pressure of the N_2 gas in the chamber was nearly the ambient pressure and was not controlled. It drifted with the ambient pressure through the measurement process. This pressure was monitored throughout the measurements, and it was corrected for in the analysis

The calibrated wavelengths from the monochromator were the wavelengths in N_2 gas. The wavelengths shown in the first column of Table 1 are wavelengths corrected to vacuum wavelengths. The wavelengths in the second column of Table 1 are wavelengths corrected to air wavelengths at 22.000 °C and a pressure of 101.325 kPa (1 standard atmosphere). To make the conversions, the well-established indices of air and N_2 gas as functions of temperature and pressure were used.[16,17]

3. RESULTS

3.1 Index of Refraction

The indices of refraction determined by the procedures discussed above are presented in Table 1. The indices are presented as relative indices in air (column 4) and absolute (vacuum) indices (column 3). The wavelengths are given as wavelengths in air (column 2) and as vacuum wavelengths (column 1). The last column gives, for each wavelength, our estimate of standard (1-sigma) uncertainty of the index. This was determined by summing in quadrature our estimates of the standard uncertainties of the various uncertainty components listed in Table 2. The total uncertainties range from near 2×10^{-5} at the shortest wavelengths, increasing to 8×10^{-5} at 14 μm. The magnitudes of these uncertainties are about what we would have expected on the basis of diffraction-dominated measurements, as discussed in the Introduction.

Table 1. Measured relative and absolute indices of refraction, and total standard uncertainties.

λ^{vac}(μm)[a]	λ^{air}(μm)[b]	Index^{vac}	Index^{air}	σ(x10^{-5})
2.2500	2.2494	4.084061	4.082973	2.5
2.5000	2.4993	4.066938	4.065855	2.2
3.0000	2.9992	4.045788	4.044711	2.1
3.3922	3.3913	4.035707	4.034633	2.0
4.0000	3.9989	4.025809	4.024738	2.2
4.5000	4.4988	4.020564	4.019494	2.3
5.0000	4.9987	4.016859	4.015790	2.4
5.5000	5.4985	4.014126	4.013058	2.6
6.0000	5.9984	4.012056	4.010988	2.7
7.0000	6.9981	4.009179	4.008112	3.0
8.0000	7.9979	4.007305	4.006238	3.9
9.0000	8.9976	4.006008	4.004942	4.3
10.0000	9.9973	4.005056	4.003990	4.8
11.0000	10.9971	4.004362	4.003296	6.4
12.0000	11.9968	4.003801	4.002736	6.9
13.0000	12.9965	4.003375	4.002310	7.5
14.0000	13.9963	4.002958	4.001893	8.0

Table 2 gives our assessment of the principal sources of uncertainty in the index results in Table 1. Column 2 gives the magnitudes of the sources, and column 3 gives the resulting contribution to the index uncertainties. When the uncertainties vary with wavelength, the uncertainties are shown as a range corresponding to wavelengths from the shortest to the longest. The largest contribution to the uncertainty for all wavelengths, except the lowest, comes from the determination of the peak angular position. This increases with wavelength, as the peak widths increase due to diffraction. The other strongly wavelength-varying uncertainty is from the radiation wavelength uncertainty of 0.2 nm, independent of wavelength. This increases rapidly as the wavelength gets shorter and is the largest uncertainty at 2.25 μm.

Table 2. Sources of uncertainty and their impacts on the index uncertainty.

Sources	Magnitude of uncertainty (1-σ)	Index uncertainty (1-σ)
Peak position	(0.8 – 5) arcseconds [1/10th of diffracted peak FWHM]	1.1 - 7.9 x 10^{-5}
Prism apex angle	0.2 arcseconds [including λ/120 (633 nm) RMS figure error]	1.2 x 10^{-5}
Index inhomogeneity	0.91 x 10^{-5} RMS for 10 mm thickness at 3.39 μm	0.91 x 10^{-5}
Encoder angle	0.2 arcseconds	0.33 - 0.34 x 10^{-5}
Prism temperature	0.005 °C	0.20 x 10^{-5}
Wavelength	0.0002 μm	1.5 - 0.01 x 10^{-5}
Alignment	0.05 arcseconds	0.08 x 10^{-5}
N_2 gas temperature	0.02 °C	0.08 x 10^{-5}
N_2 gas pressure	0.02 kPa	0.02 x 10^{-5}

3.2 Sellmeier Fit and Residuals

The air index values (at 22.000 °C) at the air wavelengths listed in Table 1 were fit by commercial least squares non-linear fitting software to several types of dispersion formulas. It was found that the best fit resulted from a standard 3-term, 6-parameter Sellmeier formula shown in Equation (2). Adding more terms or more adjustable parameters does not significantly improve the fit.

$$n^2 - 1 = \frac{K_1\lambda^2}{\lambda^2 - L_1} + \frac{K_2\lambda^2}{\lambda^2 - L_2} + \frac{K_3\lambda^2}{\lambda^2 - L_3} \quad (2)$$

The Sellmeier parameters are listed in Table 3, valid for 2.25 μm ≤ λ ≤ 14 μm. The best fit places two divergent terms (poles) below the transmission region, in the electronic absorption band wavelengths, as expected. One pole at 1.18 μm is close to the band edge, and the other at 0.40 μm is rather lower. The pole at a wavelength longer than the measured wavelengths is at 27.4 μm, in the phonon absorption band region, also not unexpected. For the formula to be consistent with the measurements within the stated uncertainties, all digits of each of the parameters in Table 3 should be used.

Table 3. Sellmeier constants of Equation (2). Range of validity: 2.25 μm < λ < 14 μm.

Sellmeier constants	
K_1	0.4886331
K_2	14.5142535
K_3	0.0091224
L_1	1.393959 μm^2
L_2	0.1626427 μm^2
L_3	752.190 μm^2

The residuals of the fit (the difference between the measured values and formula values) are shown in Figure 3. For wavelengths below 10 μm, all residuals are below 1×10^{-5}. This is consistent with lowest estimated uncertainty in this range, $\sigma = 2 \times 10^{-5}$. The largest residual in the total wavelength range is 2.3×10^{-5} at 13 μm. This is also consistent with estimated uncertainty of $\sigma = 7.5 \times 10^{-5}$ at this wavelength. Though good residual fits have little to say about systematic errors (which could be fit very well if they are smoothly varying or constant), the residuals with magnitudes at least a factor of 2 below all the estimate uncertainties, indicates that the random errors are reasonably assessed, and maybe overestimated.

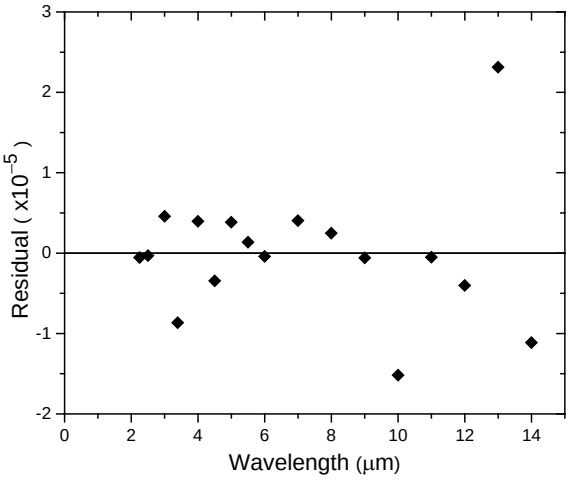


Figure 3. Plot of residuals of the fit to the data of the Sellmeier formula (2) using the parameters in Table 3.

4. COMPARISON WITH OTHER MEASUREMENTS

Comparison of the Sellmeier equation values of this work with the most extensively used data, of Salzberg *et al.*[3,4] and of Icenogle *et al.*,[5], is shown in Figure 4. The plot gives the difference between these data and our Sellmeier formula values at each of the previous author's measured wavelengths. The error bars correspond to the reported standard uncertainties for each of the measurements. The error bars for the results of this work are barely visible at the scale of the plot. The older results of Salzberg, *et al.*, are all below those of this work, by (7 to 30) × 10^{-4}, well outside their estimated uncertainty of ±3 × 10^{-4}. The Ge material they measured may well have had a substantial difference in quality, and this could account for the difference.

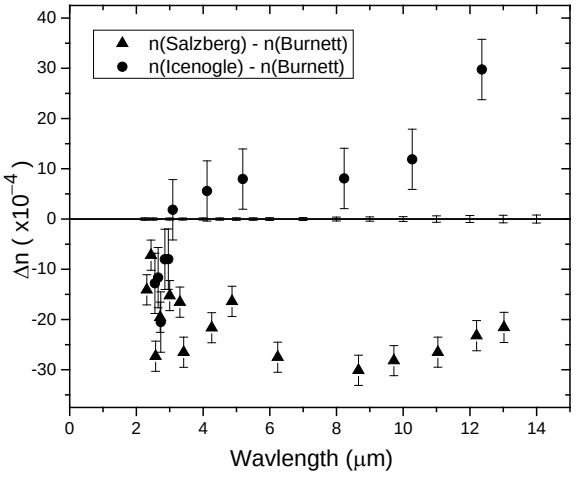


Figure 4. Plot of difference between the measurements of Salzberg *et al.*[4] and of Icenogle *et al.*,[5] and values of the Sellmeier formula of this work. The error bars represent standard uncertainties.

The newer data of Icenogle *et al.*, is closer to values of this work, with a trend of lower-than-our-values to higher-than-our-values crossing at about 2.5 μm. Their error bars include, or nearly include, our values for all wavelengths except for the value at 12.36 μm.

Figure 5 gives a similar comparison of the limited set of measurements, 8 μm to 14 μm, of Edwin *et al.*[6] to the results of this work. In this case our values are well within the error bars of Edwin, *et al*. In fact, the Edwin, *et al*. values are nearly within our error bars, which are 3 to 8 times smaller.

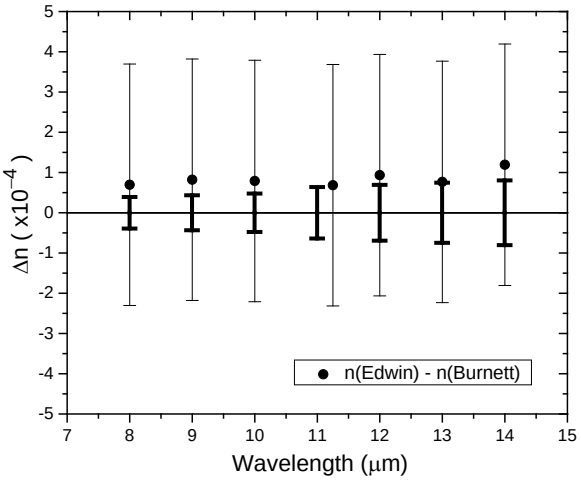


Figure 5. Plot of difference between the measurements of Edwin, *et al.*[6], and values of the Sellmeier formula of this work. The error bars represent standard uncertainties. The heavier-set error bars correspond to the results of this work.

4. CONCLUSIONS

The IR optics community has generally relied on index data for Ge measured 35 or more years ago, for optical design with Ge. It seems that these values should be revisited, especially since improvements in material fabrication may have produced higher quality materials, with potentially different index properties. We have demonstrated that we have developed a refractometry facility and procedures that are at least at the state-of-the art for index measurements in the near- to mid-IR. We have used this facility to measure the index of recently produced high-quality Ge. Our analysis of our measurement uncertainties gives values smaller than those of most previous reported measurements of Ge by about an order of magnitude or more, and our measurements approach diffraction-limited accuracy. This lowered uncertainty results from improvements of a number of the most important sources of uncertainty, including temperature, wavelength, and angle measurements. Our measurements cover the wavelength range 2.25 to 14 μm at 22.000 °C. We provide a Sellmeier formula for this range, with a standard uncertainty ranging from 2.0×10^{-5} near the short wavelength end and 8×10^{-5} at the long wavelength end.

Comparing our new Ge index data to the widely used data starting from the late 1950s, there seems to be convergence of the data over time and more consistency of the uncertainties with the differences. The remaining differences may be genuine, resulting from actual material differences, such as doping and impurity levels. A verified improvement in accuracy should be of value to assess the material differences for the new generation of Ge and other materials. We suspect that the short wavelength band edges would be particularly sensitive to defect and impurity differences, and that this would be reflected in the index behavior in this region. With the increased accuracy of our new facility, we are now measuring a number of Ge samples to explore these potential index differences.

ACKNOWLEDGEMENTS

We are pleased to acknowledge the help of Gary Wiese, of Lockheed Corporation, in helping set up our interaction with the IR optics community, helping develop the sample specifications, and getting suppliers involved in the program and getting them to provide samples.

Commercial equipment, instruments, materials, or software are identified to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology or does it imply that the items identified are necessarily the best available for the purpose.

REFERENCES

[1] Briggs, H. B., "Optical Effects in Bulk Silicon and Germanium," Phys. Rev. 77, 287 (1950).
[2] Rank, D. H., Bennett, H. E., and Cronemeyer, D. C., "Index of Refraction of Germanium Measured by an Interference Method," J. Opt. Soc. Am. 44(1), 13-16 (1954).
[3] Salzberg, C. D. and Villa, J. J., "Infrared Refractive Indexes of Silicon Germanium and Modified Selenium Glass," J. Opt. Soc. Am. 47(3), 244-246 (1957).
[4] Salzberg, C. D. and Villa, J. J., "Index of Refraction of Germanium," J. Opt. Soc. Am. 48, 579 (1958).
[5] Icenogle, H. W., Platt, B. C., and Wolfe, W. L., "Refractive indexes and temperature coefficients of germanium and silicon," Appl. Opt. 15(10), 2348-2351 (1976).
[6] Edwin, R. P., Dudermel, M. T., and Lamare, M., "Refractive index measurements of a germanium sample," Appl. Opt. 17(7), 1066-1068 (1978).
[7] Li, H. H., "Refractive Index of Silicon and Germanium and Its Wavelength and Temperature Derivatives," J. Phys. Chem. Ref. Data. 9(3), 561-658 (1980).
[8] Potter, R., "Germanium (Ge)", in Handbook of Optical Constants of Solids, Vol I., Ed. by E. Palik, Academic Press, New York, 465-478 (1985).
[9] Browder, J. S, Ballard, S. S., and Klocek, P., "Germanium, Ge", in Handbook of Infrared Optical Materials, Ed. by P. Klocek, Marcel Dekker, New York, 262-265 (1991).
[10] Daimon, M. and Masumura, A., "High-accuracy measurements of the refractive index and its temperature coefficient of calcium fluoride in a wide wavelength range from 138 to 2326 nm," Appl. Opt. 41(25), 5275-5281 (2001).
[11] Born, M. and Wolf, E, [Principles of Optics, 7th edition], Cambridge University Press, Cambridge, UK & New York, 190-193 (1999).
[12] Tentori, D. and Lerma, J. R., "Refractometry by minimum deviation: accuracy analysis," Opt. Eng. 29, 160-168 (1990).
[13] Mangum, B. W. and Furukawa, G. T., "Guidelines for Realizing the International Temperature Scale of 1990 (ITS-90)," NIST Technical Note 1265, U.S. Government Printing Office, 28-70 (1990).
[14] Schwider, J, Burow, R., Elssner, K. E., Spolaczyk, R, and Grzanna, J., "Homogeneity testing by phase shifting interferometry," Appl. Opt. 24(18), 3059-3061 (1985).
[15] Hanssen, L. M., "Integrating-sphere system and method for absolute measurement of transmittance, reflectance, and absorptance of specular samples," Appl. Opt. 40, 3196-3204 (2001).
[16] Birch, K. P. and Downs, M. J., "An Updated Edlen Equation for the Refractive Index of Air," Metrologia 30 155-162 (1993).
[17] Peck, E. R. and Khanna, B. N., "Dispersion of Nitrogen," J. Opt. Soc. Am. 56(8), 1059-1063 (1966).

Burnett, John; Kaplan, Simon.
"Refractive index measurements of Ge."
Paper presented at SPIE Optics and Photonics, San Diego, CA, United States. August 28, 2016 - September 1, 2016.

Rethinking Charge trapping and detrapping dynamic without the sheet-charge approximation

Kin P. Cheung, Jason P. Campbell and Dmitry Veksler
National Institute of Standards & Technology, Gaithersburg, MD USA, kin.cheung@nist.gov

Charge-trapping/detrapping are common occurrences that affect MOSFET performance and reliability. To understand a broad range of MOSFET phenomena, we need to think through the dynamics of charge-trapping/detrapping. The standard approach to treat oxide trapped charge is to think of them as a thin sheet of uniform charge at the centroid location [1]. The effect of this sheet charge approximation is a uniform band-bending within the oxide as well as in the semiconductor near the interface. This picture completely dominates all discussions of charge-trapping in the literature.

In the struggle to understand random telegraph noise (RTN) magnitude, the notion that the trapped charge creates a localized region of reduced (sometimes to zero) inversion charge density has gained much attention [2-8]. Physically, it makes perfect sense that the trapped charge's strong electric field greatly changes the electrostatics in its immediate surrounding. While the RTN experiments are carried out in inversion, once an electron is trapped (assuming nFET), the strong electric field changes the local electrostatic to strong accumulation (hence the absence of inversion charges). With this in mind, when treating charge emission (detrapping) one can no longer assume that it is under inversion condition as in all published work so far. Instead, it should be under strong accumulation, leading to a very different process. In this paper, we explore what this means and how to reconcile experimental observations using RTN as an example.

One way to think of the local field due to a trapped charge is to think of the peaky potential surface resulting from the atomistic random dopant simulations [9] as shown in fig.1. For nFET, the ionized dopants are negatively charged. A trapped electron in the oxide should have similar peaky potential distribution except that the trapped charges tend to be near the interface so the peak is stronger than most dopant ions.

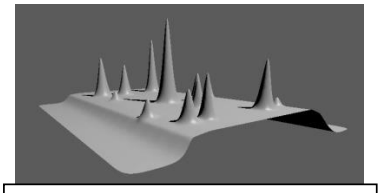


Fig. 1 Peaky surface potential in the MOSFET channel due to random dopant and electron trapped in oxide.

The peaky potentials puncture holes in the inversion layer. Within the hole, the electrostatics are very different. Fig. 2 illustrates the spatial transition from outside the trapped charge region, which is in inversion, to the center of the trapped charge region, which is in strong accumulation.

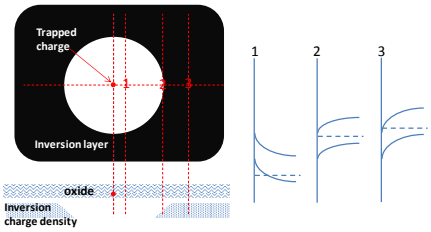


Fig. 2 A cored-out hole in the inversion layer surrounds the trapped electron. Three regions are identified to show the corresponding band bending at the bottom of the figure.

Fig. 3 illustrates the difference in band diagrams between the conventional sheet charge model and the localized charge model. A trap site inside the oxide captures a charge from inversion layer through tunneling (right hand side). Upon charge capture the trap energy reduces through multi-phonon relaxation [10]. Two sets of experiments in the literature both suggest that this energy drop is ~1.5 eV [11, 12]. Thus the new trap energy is shown to be below the valence band edge. The top panel of left hand side is the usual sheet charge picture whereas the bottom panel of left hand side is the local charge picture. It should very obvious that the emission process of these two cases are very different.

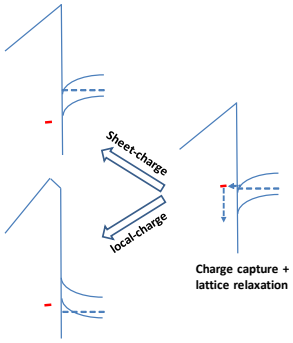


Fig. 3 Band diagram showing the capture of an electron by a defect site inside the oxide (RHS). The resulting band diagram for the conventional sheet charge model (top LHS) and the new local charge model (bottom LHS) are quite different.

We note that once multi-phonon relaxation is included, the emission kinetics are very difficult to explain using the conventional sheet charge model. The trapped charge must be thermally excited to overcome a large energy

Campbell, Jason; Cheung, Kin; Veksler, Dmitry.
"Rethinking Charge trapping and detrapping dynamic without the sheet-charge approximation."
Paper presented at 47th IEEE Semiconductor Interface Specialists Conference, San Diego, CA, United States. December 8, 2016 - December 10, 2016.

barrier of ~ 1.5 eV. The probability is too low. Even if emission occurs to the gate electrode, there is still a large energy barrier. This is one of the reasons that RTN kinetics are difficult to model. Indeed, Kirton and Uren *et al*. are forced to conclude that the relaxation energy must be in the 20 to 150 meV range only [13]. Similarly, Palma *et al*. also relies on this same shallow relaxation to explain the RTN emission time constant [14]. These assumptions are, of course, not in agreement with experimental evidences [11, 12].

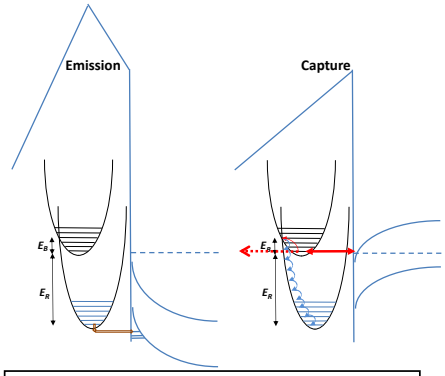


Fig. 4 Detailed charge capture and emission band diagram in the local charge model including multi-phonon relaxation.

With the local charge model, the opposite becomes true. The band diagram charge capture and emission in the local charge model is illustrated in more detail in fig.4. As can be seen, emission becomes very efficient. Capturing a hole from the accumulation layer is energetically extremely favorable. This would lead to very short emission time rendering RTN unobservable. This dilemma is resolved when we once again consider the highly localized nature of the field, leading to strong quantum confinement effect. This quantization is much stronger than the familiar inversion layer quantum confinement effect because the local accumulation layer is strongly confined in all three dimensions. Fig. 4 already shows the quantized energy levels in the accumulation layer.

Tunneling rate depends on two factors. One is the barrier (height and width), the other is the density of states (both the available carrier for tunneling and the available states to tunnel to). There are very few states in the accumulated region and that the first state is at least a fraction of an eV from the top of the conduction band. Note that the reduction in available states will cause the band bending to respond to the vertical field and become stronger, further reducing the density of states and increasing the distance of the first state from the top of the valence band. Thus even though the electron is at energy below the valence band edge, the emission still requires phonon assisted tunneling to reach the first state as illustrated in figure 4. Coupled with the low density of state available the emission rate is low enough for RTN to be observable in a reasonable measurement window.

The capture and emission process illustrated in figure 4 is consistent with the temperature effect on capture and emission time constants – both are thermally activated but with different activation energies [15].

Experimentally, the majority of the reported RTN time constants have the following behavior: as gate bias decreases, capture time constant increases and emission time constant decreases [15]. The model in figure 4 explains the gate bias dependent capture time constant easily, but what happen to the emission time constant? When gate overdrive reduces, quantum confinement also relaxes. The result is an increase in number of accessible states and a reduction in the energy offset of the first state from the valence band edge. Both effects will increase the emission rate and reduce the emission time constant – in agreement with majority of data in the literature.

[1] Wolters, D. R. and J. J. van der Schoot, *J. Appl. Phys*. **58**(2): 831-837(1985); See also Taur, Y. and T. H. Ning, *Fundamentals of Modern VLSI Devices,* page 87. Cambridge University Press, 1998. ISBN 0-521-55959-6. [2] L. D. Yau, and C.-T. Sah, *IEEE Trans. Electron Dev.*, **16**(2): 170-177(1969). [3] G. Reimbold, *IEEE Trans. Electron Dev.*, **31**(9): 1190-1198(1984). [4] A. Ohata, A. Toriumi, *et al., J. Appl. Phys.* **68**(1): 200-204(1990). [5] E. Simoen, B. Dierickx, *et al., IEEE Trans. Elect. Dev.,* **39**(2), 422(1992). [6] A. Asenov, R. Balasubramaniam, *et al., IEEE Trans. Elect. Dev.,* **50**(3): 839(2003). [7] K. Sonoda, K. Ishikawa, *et al., IEEE Trans. Elect. Dev.,* **54**(8): 1918(2007). [8] Southwick, R. G., K. P. Cheung, *et al., IEEE Silicon Nanoelectronics Workshop* (SNW), Kyoto, Japan, 2012. [9] A. Asenov *et al., J. Comput. Electron*., **54**(8), 349-373(2009). [10] Henry, C. H. and D. V. Lang, *Physical Review B* **15**(2): 989 LP -1016(1977). [11] Lu, Y. and C. T. Sah, *J. Appl. Phys*. **78**(5), 3156-3159(1995). [12] Takagi, S., N. Yasuda, *et al., Int. Electron Devices Meeting*, (IEDM) 1996, pp323-326. [13] Kirton, M. J. and M. J. Uren, *Appl. Phys. Lett*. **48**(19), 1270-1272(1986). [14] Palma, A., A. Godoy, *et al. Phys. Rev. B* **56**(15), 9565-9574(1997). [15] Ralls, K. S., W. J. Skocpol, *et al, Phys. Rev. Lett.* **52**(3), 228 LP-231. (1984).

Campbell, Jason; Cheung, Kin; Veksler, Dmitry.
"Rethinking Charge trapping and detrapping dynamic without the sheet-charge approximation."
Paper presented at 47th IEEE Semiconductor Interface Specialists Conference, San Diego, CA, United States. December 8, 2016 - December 10, 2016.

Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs*#

Christopher E Sunday, Dmitry Veksler, Kin C. Cheung, Yaw S. Obeng
Engineering Physics Division,
Physical Measurement Laboratory,
National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, MD 20899

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

Traditional metrology has been unable to adequately address the needs of emerging integrated circuits at the nano scale; thus, new metrology and techniques are needed. For example, the reliability challenges, and their root causes, in TSV enabled 3D-IC fabrication need to be well understood and controlled to enable mass production. This requires new approaches to the metrology. In this paper, we use microwave propagation characteristics (S-parameters) to study the thermo-mechanical reliability issues that precede electromigration failures in Cu-filled TSVs. The pre-failure insertion losses and group delay are reversibly dependent on both the device temperature and the amount of current forced through the devices under test. This is attributed to the plasticity of the copper fill in the TSVs.

Introduction

Metrology is needed to characterize, at the nano scale, the structure and composition in emerging integrated circuits. The reliability of the interconnects in such systems is a critical issue; for example, at 1 THz, the skin depth of Cu is only 65 nm, and even if the wire diameter remains several microns in size for 3-D interconnects, the current carrying capacity of such interconnects will be severely limited. In fact, the introduction of TSV enabled 3D-IC to the market has been hindered by many reliability challenges. Particularly, thermal stress related failures in this new technology are significantly worse than in the traditional planar integrated circuits. The stress buildup, due to mismatch in the thermal properties of the materials of construction, results in generation of defects such as cracks, voids, delamination, plastic deformation, substrate warping and buckling. This, in turn, introduces new reliability challenges such as the degradation of transistors in close proximity to the TSVs[1]. This situation is not all that unusual; about

Cheung, Kin; Obeng, Yaw; Sunday, Christopher; Veksler, Dmitry.
"Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs."
Paper presented at 231st ECS MEETING, New Orleans, LA, United States. May 28, 2017 - June 2, 2017.

65% of all microelectronic device failures are thermo-mechanical related, mostly due to thermally induced stresses and strains stemming from the mismatch in the neighboring dissimilar materials[2]. Thus, the various degradation processes and their impact on the different components have to be evaluated and well understood to enable mass production.

We have shown elsewhere that the microwave scattering characteristics in interconnects are sensitive to defect formation and material transformation[3, 4], and should be sensitive to the thermomechanical dynamics in electromigration (EM). EM is essentially the flux of material carried along with the dense electron current, that leads to stress build-up and void formation resulting in degraded device performance[5]. Given the large volume via fill in TSVs, the change in low frequency resistance (R_{dc}) from the void formation is difficult to measure, and may require the use of high temperatures and large forced currents to induce measurable damage. Indeed, during EM studies on TSV-enabled samples R_{dc} measurably increased only when voids larger than the TSV conductive diameter formed, forcing the electron flow to go through the resistive barrier shunts[6]. Thus, by the time a measurable resistance is detected the device under test (DUT) would have been long destroyed. On the other hand, the onset of void formation results in impedance changes which are easily measured with the insertion losses of broadband microwave spectrum. Furthermore, the phase changes in the propagating microwave can yield additional mechanistic information, such as changes in dielectric properties of the materials of construction.

In this paper, we discuss application of broadband microwave spectroscopy to investigate the pre-failure thermo-mechanical issues of TSV enabled 3D ICs during EM. Specifically, we present the results of a preliminary demonstration on the use microwave propagation characteristics (S-parameters) to study the impact of temperature and current on the pre-catastrophic-damage and thermo-mechanical reliability issues during EM of Cu TSV-based interconnects in 3D-ICs.

Experiments

Otherwise identical devices with different thermal histories were compared at various elevated temperatures, and direct currents, while monitoring the microwave insertion losses. Three types of DUT thermal histories were studied:

- Group 1: As received devices that had been stored in a dry N_2-box for about 18 months (green markers in Figures 1, 3 and 4, respectively);
- Group 2: The group 1 devices that had been subjected to up to 500 thermal cycles in the 30ºC to 125ºC temperature range (red markers in Figures 1, 3 and 4, respectively)
- Group 3: The group 1 devices that had been stored at 200C under N_2 for 72 hours, and then stored at room temperature for another 72 hours ((black markers in Figures 1, 3 and 4, respectively).

This initial demonstration involved about three of each device type. The EM-studies experimental plan is summarized in Table 1.

Cheung, Kin; Obeng, Yaw; Sunday, Christopher; Veksler, Dmitry.
"Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs."
Paper presented at 231st ECS MEETING, New Orleans, LA, United States. May 28, 2017 - June 2, 2017.

Dedicated group-signal-ground (GSG) test structures were used in these experiments. TSV-enabled two-level stacked dies, bonded together with polymer were used. The TSVs were located in the top chip. The samples were placed on a heated chuck in an open laboratory ambient, and dc current was forced through the signal line. The microwave signals were intermittently introduced and monitored with a 2-port network analyzer once the DUT reached the target temperature. Single devices from each group were subjected to step-wise increase in forced current at specific device temperatures. RF probes were landed on the GSG bond pads after the device reached the desired temperature to minimize drifts from expansion mismatch and potential damage to the probe tips.

Table 1: Compendium of materials and test conditions uses in the electromigration study.

Thermal History	EM Temperatures / ºC	Forced DC Current Range / mA
Group 1	25, 125	0-300
Group 2	25, 125	0-250
Group 3	300	0-250

For each sample, maximum current was maintained for up to 72 hours, with intermittent microwave measurements. At the end of the desired hold time, both the forced dc current and DUT heating sources were turned off. Insertion losses were measured again after a maximum of 72 hours at room temperature, to check DUT recovery.

Results and Discussion

Figure 1 shows the insertion-loss (S21) magnitude, at 35 MHz, as a function of the temperature and the forced dc current during the EM test. Relatively large changes in the insertion losses (S21 amplitude) were observed only at the highest temperature studied (300°C). As shown in Figure 2, the S21 losses were reversible. Specifically, the S21 losses were recovered once the samples were cooled back to room temperature, irrespective of thermal history and the maximum temperature or current endured.

As shown in Figure 2, the S21 amplitude depends on both the chuck temperature and forced dc current, and is reversible. The insertion losses increased with increasing DUT temperature and increasing dc current. Once the heating and forced currents were turned off the S21 returned to close to the original values measured at the start of the experiments.

The thermodynamics of the insertion losses were investigated; the Arrhenius-type dependence (as shown in Figure 3) suggests that the insertion losses are related to reversible temperature driven processes within the interconnect system. Interestingly, while the pre-activation factor did not change, the slopes (i.e., the activation energies) of the Group1 and Group 2 devices are different; the Group 3 devices were not tested. This suggests that the details of the RF signal loss may be dependent on the thermal history, i.e., stress state of the interconnects in the device under test[7, 8]. While material transport during EM in copper interconnect systems is known to occur predominantly along the Cu / capping layer interfaces, [9] grain boundary diffusion plays a significant role[10]. The reversibility of insertion losses suggest temperature-driven structural / phase

Cheung, Kin; Obeng, Yaw; Sunday, Christopher; Veksler, Dmitry.
"Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs."
Paper presented at 231st ECS MEETING, New Orleans, LA, United States. May 28, 2017 - June 2, 2017.

transformations at the copper-dielectric interface and / or at copper grain boundaries[11, 12], that could lead to EM-induced plasticity in the copper interconnects[13]

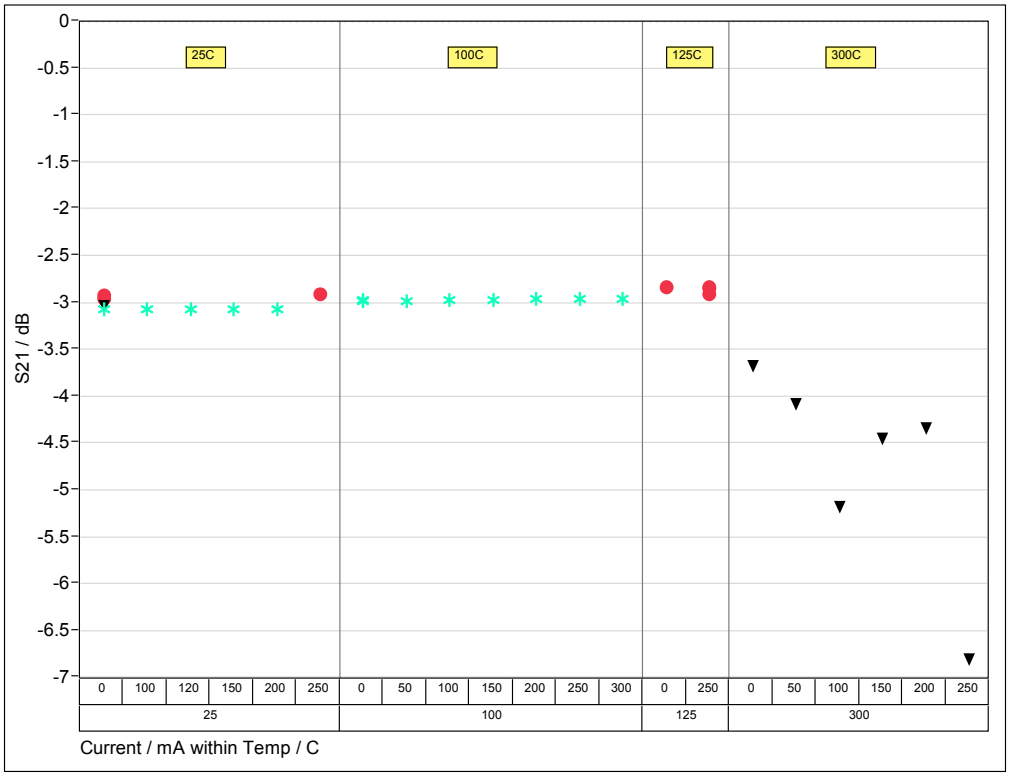


Figure 1: Impact of forced dc current (from 0 to 300 mA, as shown in the first row of the table) at various temperature 25°C, 100 °C, 125 °C and 300 °C, respectively, on microwave insertion loss (S21/dB) at 35 MHz

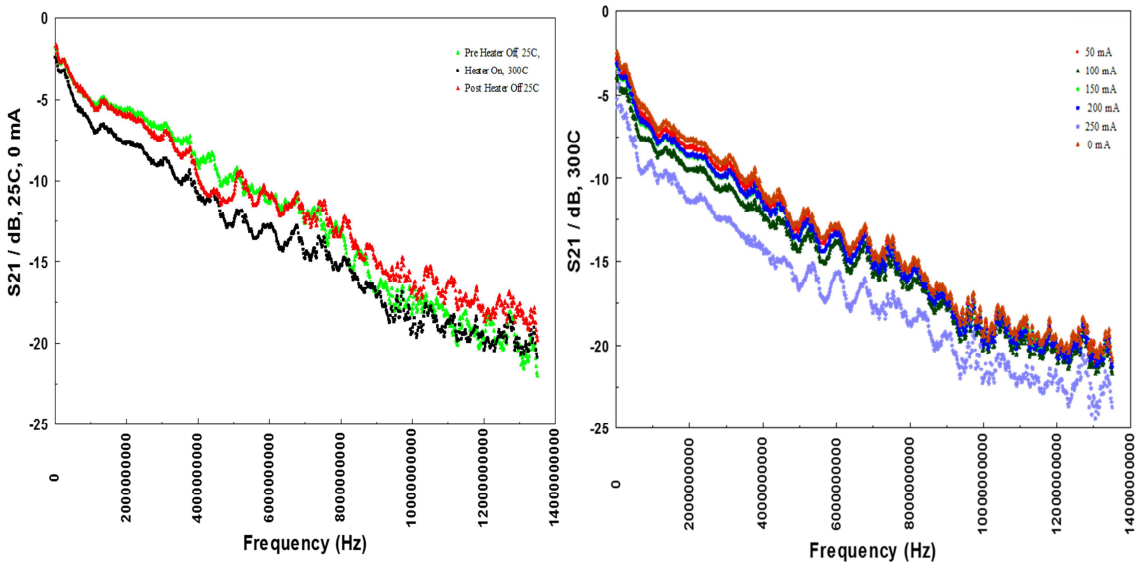


Figure 2A: Recoverable changes in S21 as a function DUT temperature Figure 2B S21 changes as a function of forced dc current at 300ºC for a typical group-3 device

The microwave propagation speed through the DUT is determined by the dielectric properties of the materials of construction[12]. Group delay is the slope of the phase angle dependence on the microwave frequency. Figure 4 shows the dependence of the microwave group delay as a function of forced dc current for all the devices studied, irrespective of thermal history or test temperature. The changing phase angle depends on the dielectric properties of the DUT; thus, the dielectric permittivity of the DUT depend on the magnitude of the forced dc current. The data suggest that the forced dc-current induces localized Joule heating in the interconnects, and increases the effective dielectric constant of the DUT.

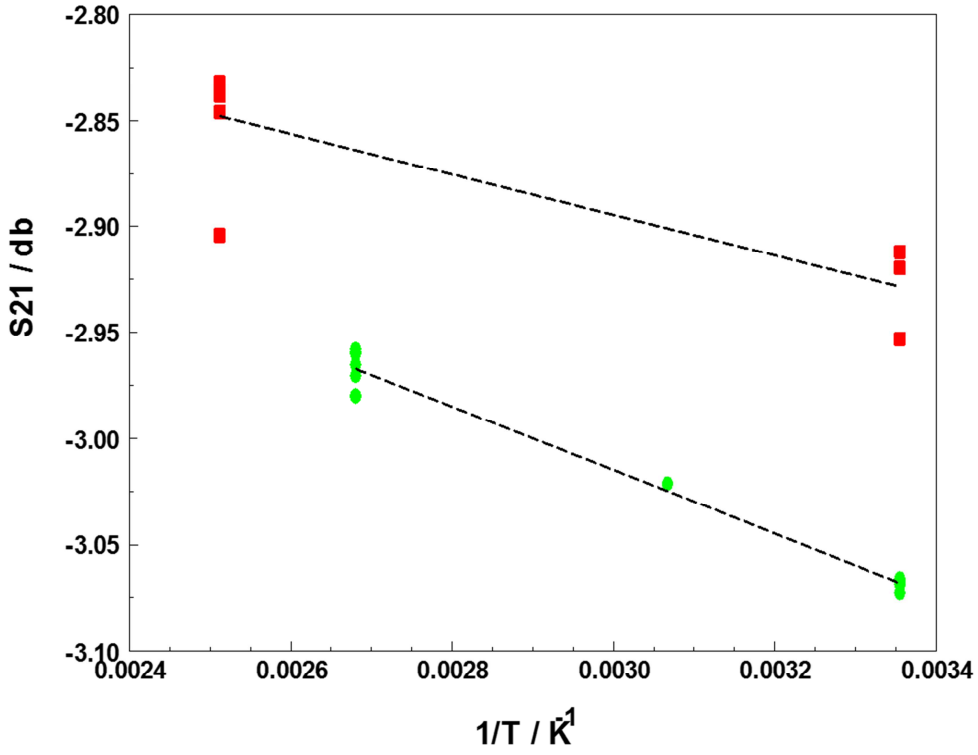


Figure 3: Impact of device temperature on microwave insertion loss (S21/dB) on device Group 1 (green dots) and Group 2 (red dots), respectively, monitored at 35 MHz.

Cheung, Kin; Obeng, Yaw; Sunday, Christopher; Veksler, Dmitry.
"Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs."
Paper presented at 231st ECS MEETING, New Orleans, LA, United States. May 28, 2017 - June 2, 2017.

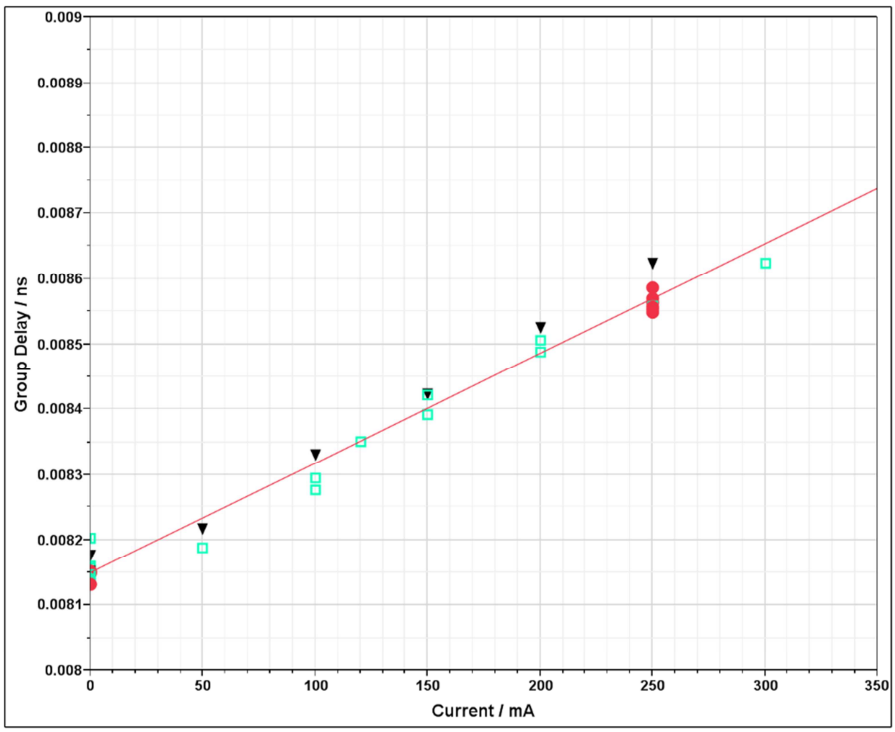


Figure 4: Impact of dc current at on group delay during microwave propagation at 35 MHz

Conclusions

Impedance changes occur long before electromigration induced damage in TSV enabled 3D-ICs interconnects. These changes are reversible in the low forced current domain, and on the short stress time-scales. Pre-existing mechanical damage in the interconnects, do not appear to affect the reversibility of the impedance changes.

Cheung, Kin; Obeng, Yaw; Sunday, Christopher; Veksler, Dmitry.
"Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs."
Paper presented at 231st ECS MEETING, New Orleans, LA, United States. May 28, 2017 - June 2, 2017.

References

1. A. Mercha, G. Van der Plas, V. Moroz, I. de Wolf, P. Asimakopoulos, N. Minas, S. Domae, D. Perry, M. Choi, A. Redolfi, C. Okoro, Y. Yang, J. Van Olmen, S. Thangaraju, D. Sabuncuoglu Tezcan, P. Soussan, J. H. Cho, A. Yakovlev, P. Marchal, Y. Travaly, E. Beyne, S. Biesemans and B. Swinnen, presented at the Electron Devices Meeting (IEDM), 2010 IEEE International, 2010 (unpublished).
2. B. Wunderle and B. Michel, Microelectronics Reliability **46** (9), 1685-1694 (2006).
3. C. Okoro, P. Kabos, J. Obrzut, K. Hummler and Y. S. Obeng, Electron Devices, IEEE Transactions on **60** (6), 2015-2021 (2013).
4. C. Okoro, J. W. Lau, F. Golshany, K. Hummler and Y. S. Obeng, Electron Devices, IEEE Transactions on **61** (1), 15-22 (2014).
5. J. Huang, A. Oates, Y. Obeng and W. L. Brown, Journal of The Electrochemical Society **147** (10), 3840-3844 (2000).
6. T. Frank, S. Moreau, C. Chappaz, P. Leduc, L. Arnaud, A. Thuaire, E. Chery, F. Lorut, L. Anghel and G. Poupon, Microelectronics Reliability **53** (1), 17-29 (2013).
7. C. Okoro, L. E. Levine, R. Xu, K. Hummler and Y. S. Obeng, IEEE Transactions on Electron Devices **61** (7), 2473-2479 (2014).
8. C. Okoro, L. E. Levine, R. Xu, K. Hummler and Y. Obeng, Journal of Applied Physics **115** (24), - (2014).
9. L. Zhang, P. S. Ho, O. Aubel, C. Hennesthal and E. Zschech, Journal of Materials Research **26** (21), 2757-2760 (2011).
10. L. Arnaud, T. Berger and G. Reimbold, Journal of Applied Physics **93** (1), 192-204 (2003).
11. Y. Mishin, M. Asta and J. Li, Acta Materialia **58** (4), 1117-1151 (2010).
12. T. Frolov, S. V. Divinski, M. Asta and Y. Mishin, Physical Review Letters **110** (25), 255502 (2013).
13. A. S. Budiman, C. S. Hau-Riege, W. C. Baek, C. Lor, A. Huang, H. S. Kim, G. Neubauer, J. Pak, P. R. Besser and W. D. Nix, Journal of Electronic Materials **39** (11), 2483-2488 (2010).

Cheung, Kin; Obeng, Yaw; Sunday, Christopher; Veksler, Dmitry.
"Understanding the Pre-Failure Thermo-Mechanical Issues In Electromigration of TSV Enabled 3D ICs."
Paper presented at 231st ECS MEETING, New Orleans, LA, United States. May 28, 2017 - June 2, 2017.

Multiheterodyne Spectroscopy Using Multi-frequency Combs

David F. Plusquellic[1], Gerd A. Wagner[1], Adam J. Fleisher[2], David A. Long[2] and Joseph T. Hodges[2]

[1]Physical Measurement Laboratory, National Institute of Standards and Technology, Boulder, CO 80305
[3]Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899
david.plusquellic@nist.gov

Abstract: Near-IR dual frequency combs generated from waveform driven electro-optic phase modulators (EOMs) are used for high resolution studies in low pressure cells and for remote sensing from natural targets (Boulder Flatirons). Arbitrary waveform generators (with up to 20 GHz of bandwidth) enable amplitude (AM), phase (PM) and harmonic (HM) modulation control of dual parallel phase modulators to generate tailored frequency combs for sensing requirements.
OCIS codes: (280.0280) Remote Sensing and Sensors; (300.6310) Spectroscopy, heterodyne

Frequency combs generated using radiofrequency(rf)-driven electro-optic phase modulators have proved advantageous in a wide variety of areas from low pressure gas cell studies [1] to sub-Doppler spectroscopy [2]. Dual comb methods [3-7] are particularly valuable to enable high throughput collection over wide spectral regions that exceed detector bandwidths for a single comb. Dual parallel Mach-Zehnder phase modulators (MZM) driven with arbitrary waveform generators (AWGs) bring additional real time single element control of the interferogram to optimize comb properties. Comb tooth resolution/interferogram scan rate, optical bandwidth and relative phase can be readily adapted to sensing requirements while maintaining optimal conditions for data throughput.

In this work, we generate frequency combs using MZMs driven by AWGs that originate from a single external cavity diode laser tunable over a range from 1600 nm to 1646 nm. The basic elements of the system are shown in Fig. 1 (left) [6]. For the gas cell experiment, the output of the ECDL is fiber split to generate combs in two MZMs for the probe (+reference) and local oscillator (LO) arms. The signal arm is split again and separate frequency offsets are added using AOMs to provide probe and reference combs for spectral normalization. The probe (after passing through the gas cell) and reference combs are combined and then mixed with the LO comb. Multiheterodyne signals are detected on a 700 MHz NIR detector and digitized at rates up to 3 GHz depending on the comb spectral bandwidth. For the remote sensing experiments, scatter from the Rocky Mountain Flatiron at a distance of ≈2.8 km is detected using a 28 cm telescope and photomultiplier tube for photon counting. In this case, the separate reference arm is not used. The signal and LO combs are combined and amplified to 13 mW before beam expansion and transmission to the atmosphere [8]. A small portion to generate the reference comb is fiber split after the final amplifier. In all cases, the data are streamed to memory with bandwidths up to 1.6 mega-sample/sec and processed and averaged in real time using a 56 hyper-threaded core computer before saving to disk.

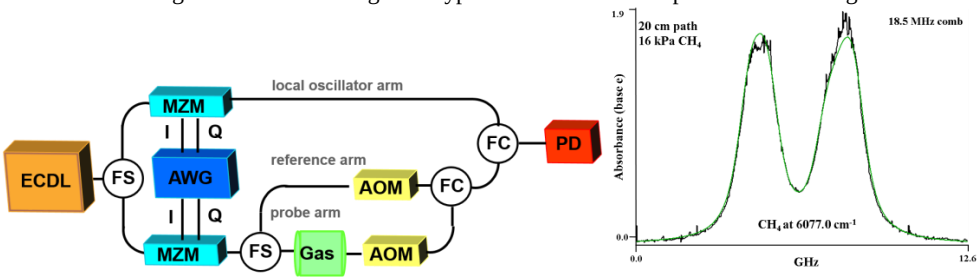


Fig. 1. Multiheterodyne spectrometer used for low pressure gas and remote sensing studies. ECDL, external cavity diode laser; FS, fiber splitter; FC: fiber coupler; MZM: dual parallel Mach Zehnder electro-optical phase modulator with in-phase (I) and quadrature (Q) inputs; AWG: arbitrary waveform generator; AOM: acousto-optical modulator; PD: photo-diode. Multiheterodyne spectrum of CH_4 (right) consisting of the average of 2 independent combs each normalized to its corresponding reference comb.

For the gas cell work, three frequency combs were produced using AM+PM waveforms. The waveform frequencies of first comb and second comb were 666.66667 MHz and 666.61667 MHz, respectively. The AM period of probe signal (third comb) is 54 ns to generate fine teeth with spacing 18.518 MHz and the period of the second comb (LO comb) is the same as the sampling period of 2.5 ms. The AM/PM combs were optimized (in an iterative way using genetic algorithms) by variation of the amplitudes, phases and pulse widths/shape of the fundamental and

Fleisher, Adam; Hodges, Joseph; Long, David; Plusquellic, David; Wagner, Gerd.
"Multiheterodyne Spectroscopy Using Multi-frequency Combs."
Paper presented at CLEO, San Jose, CA, United States. May 14, 2017 - May 19, 2017.

two harmonics over the sampling period. The digitized heterodyne waveform is Fourier transformed to give independent sets of rf combs for each of the probe and reference arms. A secondary set is also within the detection bandwidth but is reversed and offset relative to the first set. Each of the four combs contain nearly 700 teeth covering a 13.3 GHz region. For each set, the probe and reference combs are sorted, integrated and then normalized to the corresponding reference signals.

The amplitude transmission versus detuning of a multiheterodyne spectrum of CH_4 at 6077 cm^{-1} is shown in Fig. 1 (right). The spectrum was obtained using 20 cm long sample cell filled with 16 kPa of methane at room temperature and was acquired in 40 ms. The integrated intensity and doublet structure are in good agreement with predictions from the HITRAN database shown superimposed.

For the remote sensing study of CO_2 at 6241.40 cm^{-1}, the interoferograms of the probe (PMT) and reference signals are shown in Fig. 2 after time shifting the coadded probe interoferograms for the round trip transit time. The AOM frequency was 50 MHz to minimize the incoherence from the scattering. The MZM were driven at 1.000 GHz and 1.001 GHz giving an interferogram scan rate of 1 MHz to minimize effects of atmospheric turbulence. The data were acquired in <8 mins at an average count rate of $>10^6$ counts/sec. Phase jitter and drift in the system were corrected in real time using cross-correlations of the reference data with initial saved data. These same shifts were applied to the binning of the photon counts during accumulation. While the CO_2 fractional absorption is approximately correct, the large scatter that still remains is likely associated with reference channel normalization and wavelength decorrelation issues which are being addressed using a random distribution of relative phase shifts and stepped motion of the transceiver.

AWG-driven MZM generated combs enable agile changes in the comb spacing and bandwidth across a wide range (<1 MHz to >20 GHz) and therefore increase the utility of these methods for high resolution applications.

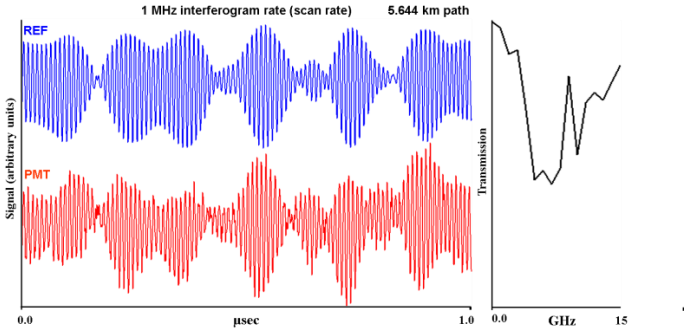


Fig. 2 10^8 averaged reference (top left) and coadded photon counting (bottom left) interferograms of CO_2/H_2O near 6241.4 cm^{-1} over 8 mins (20% throughput) acquired at 1 GHz sampling rate (left) and the normalized transmission spectrum (right) with fifteen 1 GHz comb teeth.

1. P. Martín-Mateos, M. Ruiz-Llata, J. Posada-Roman, and P. Acedo, "Dual comb architecture for fast spectroscopic measurements and spectral characterization," IEEE Photonics Technol. Lett. **27**(12), 1309–1312 (2015).
2. D. A. Long, A. J. Fleisher, D. F. Plusquellic, and J. T. Hodges, "Multiplexed sub-Doppler spectroscopy with an optical frequency comb," Phy. Rev. A. **94**, 061801(1-4) (R) (2016).
3. F. Keilmann, C. Gohle, and R. Holzwarth, "Time-domain mid-infrared frequency-comb spectrometer," Opt. Lett. **29**, 1542-1544 (2004).
4. I. Coddington, W. C. Swann, and N. R. Newbury, "Coherent multiheterodyne spectroscopy using stabilized optical frequency combs," Phys. Rev. Lett. **100**(1), 013902 (2008).
5. V. Torres-Company and A. M. Weiner, "Optical frequency comb technology for ultra-broadband radiofrequency photonics," Laser Photonics Rev. **8**(3), 368–393 (2014).
6. D. A. Long, A. J. Fleisher, K. O. Douglass, S. E. Maxwell, K. Bielska, J. T. Hodges, and D. F. Plusquellic, "Multiheterodyne spectroscopy with optical frequency combs generated from a continuous-wave laser," Opt. Lett. **39**(9), 2688–2690 (2014).
7. A. J. Fleisher, D. A. Long, Z. D. Reed, J. T. Hodges and D. F. Plusquellic, "Coherent cavity-enhanced dual-comb spectroscopy", Opt. Express **24**, 10424 (2016).
8. G. A. Wagner, and D. F. Plusquellic, "Ground-based, integrated path differential absorption LIDAR measurement of CO_2, CH_4 and H_2O near 1.6 µm Region", *Appl. Opt.* **55(23)**, 6292-6310 (2016).

Fleisher, Adam; Hodges, Joseph; Long, David; Plusquellic, David; Wagner, Gerd.
"Multiheterodyne Spectroscopy Using Multi-frequency Combs."
Paper presented at CLEO, San Jose, CA, United States. May 14, 2017 - May 19, 2017.

A Focal Plane Imager with High Dynamic Range to Identify Fabrication Errors in Diffractive Optics

M. Imran Afzal, Sofia C. Corzo-Garcia, Ulf Griesmann
National Institute of Standards and Technology (NIST), Engineering Physics Division, Gaithersburg, MD 20899-8220, USA
muhammad.afzal@nist.gov

Abstract: We describe the prototype of a focal plane imager for the characterization of fabrication errors in diffractive optics. We also demonstrate the use of a high dynamic range imaging method to identify small errors in a hologram and a grating that can be traced to the fabrication method.
OCIS codes: (050.0050) Diffraction and gratings; (090.0090) Holography; (120.3930) Metrological instrumentation; (120.3940) Metrology

1. Introduction

All micro-fabrication methods used in the fabrication of diffractive optics impart characteristic fabrication errors to the diffractive optic. Especially when diffractive optics are used in demanding imaging applications it is important that those fabrication errors are carefully characterized, well understood, and eliminated. Otherwise fabrication errors result in light being scattered in unwanted directions causing flare and reduced contrast, or even ghost images. The metrology methods most familiar to optics fabricators are not best suited to the evaluation of fabrication errors in diffractive optics. Optical interferometers are only useful for measurements in a small spatial frequency range near zero, and microscope-based inspection systems have a very limited field of view that is typically much smaller than the diffractive surface. Here we describe a prototype measurement system that analyzes the light reflected by a diffraction grating or diffractive optic in the focal (Fourier) plane of a lens, with the goal to separate the small amount of stray light caused by fabrication errors from the strong signal of the light diffracted at the design angle. This idea is not new. A focal plane scanner was used many years ago to evaluate the intensities of ghost lines present in diffraction gratings produced at the Mount Wilson Observatory [1]. A similar focal plane scanner using contemporary technology was described recently for the evaluation of errors in gratings that were fabricated with e-beam lithography [2].

These focal plane scanners have an important limitation. While they can achieve the necessary dynamic range, the measurements are only one-dimensional, the signal at the focal plane of a focusing lens is inherently two-dimensional. The prototype focal plane imager we describe here is designed to overcome this limitation of the focal plane scanners. The light reflected by a grating or the sub-aperture of a diffractive optic is focused on a charge coupled device (CCD) sensor using a lens. An additional innovation is our use of a high dynamic range (HDR) imaging method to achieve the required high dynamic range. We describe the design of the focal plane imager and illustrate its performance – and its current limitations – with measurements of three different gratings.

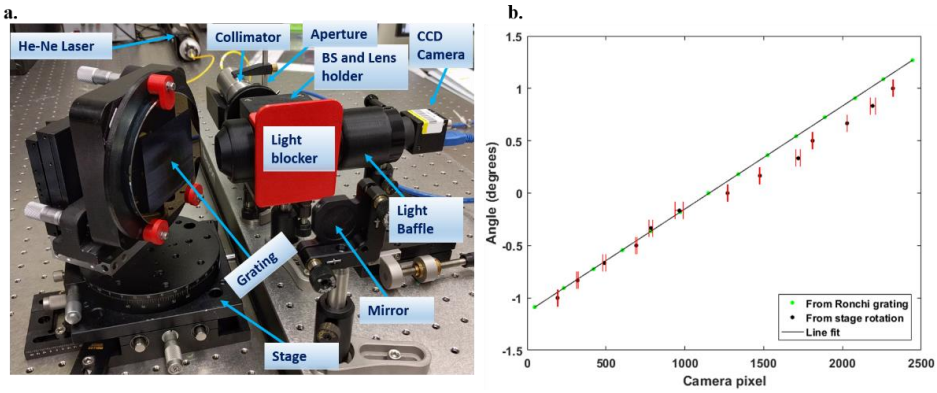


Figure 1. Measurement setup (a) and diffraction angle calibration of the camera (b). The mean of the detector position repeatabilities indicated by the error bars is 7.9 pixels (1σ). The jump in the stage angle data is due to backlash.

Afzal, Muhammad; Corzo Garcia, Sofia; Griesmann, Ulf.
"A Focal Plane Imager with High Dynamic Range to Identify Fabrication Errors in Diffractive Optics."
Paper presented at Optical Fabrication and Testing, Denver, CO, United States. July 9, 2017 - July 13, 2017.

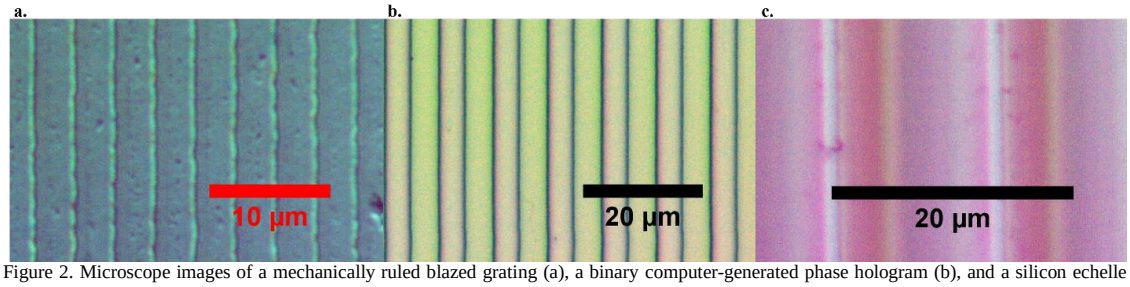


Figure 2. Microscope images of a mechanically ruled blazed grating (a), a binary computer-generated phase hologram (b), and a silicon echelle grating (c).

2. Measurement setup

The measurement setup is shown in Fig. 1a. Light from a frequency stabilized helium-neon (He-Ne) laser operating at 632.8 nm is transferred from the laser to a laser collimator by a single mode optical fiber. It is important that the laser collimator produces a well-apodized Gaussian beam that is not diffracted at the collimator. The measurements described here were made with a commercial laser collimator that has an exit aperture of 24 mm diameter and the beam diameter of the Gaussian beam at the $1/e^2$ point is approximately 11 mm. No diffraction effects due to the collimator ("ringing") were observed even in the measurements with the highest dynamic range. The collimator is followed by an iris aperture that remained fully open during measurements. The iris is only used to create a narrow beam for alignment purposes. A beam splitter (BS in Fig. 1) reflects 50% of the light from the laser collimator at a right angle onto the grating under test. The test gratings are mounted on a simple manual 5-axis mount that can be seen to the left in Fig. 1a. The main feature of the mount is a rotation stage that is used to rotate the grating until the Littrow condition is met and the test beam is retro-reflected by the grating for the desired diffraction order. Light reflected by the test grating traverses the beam splitter and is focused by a plano-convex lens with 100 mm focal length onto the CCD image detector in the camera. The second beam exiting the beam splitter is blocked by an absorber (attached to the red holder in Fig. 1a) in the measurements described here. We found the design of the beam splitter to be critical for measurements with low stray light level. The measurements described in the following section were made with a cube beam splitter with 25.4 mm edge length. A cube beam splitter is not ideal because the cube faces, even with good anti-reflection (AR) coating, cause substantial stray reflections that reach the camera. A custom beam splitter is being fabricated, but it was not available at the time the measurements described here were made. The camera has an uncooled 2/3 inch CCD sensor with 2448 x 2048 pixels and 12 bits resolution (6.5 bits noise). The window of the camera housing and the cover glass on the image sensor were both removed to eliminate interference and stray light. The camera is mounted on a translation stage to enable us to position it precisely at the focus of the lens. Light baffles reduce the amount of stray light reaching the camera and during measurements the whole setup was enclosed in a box that was made from black foam board and sealed with black caulk.

Both gain and exposure time of the camera can be programmed in a wide range. Images were acquired for 18 settings of gain and exposure time with increasing sensitivity. For each of the settings 10 images were acquired and averaged. Using the gain and exposure time settings the images were scaled onto a common intensity scale and were averaged after removing saturated pixels and the noise floor from each of the 18 images. The result is a single HDR image with a dynamic range of about 10^7. The calibration of the lateral image scale in terms of the (relative) diffraction angle was accomplished in two ways. A mirror was mounted in the test mount and the position of the reflection on the image detector was determined. The mirror was then rotated by small angle increments that could be measured using a Vernier angle scale on the rotation stage and the reflection position was noted for each of the angles. The result is shown in Fig. 1b. The error bars indicate the position repeatability (1σ) for five repeat measurements. The angle scale was also calibrated using a chrome-on-glass Ronchi grating with a period of 100 µm, which creates 14 diffraction peaks across the image detector. The known angular separation of the peaks is then used to establish the angle scale and the result is also shown in Fig. 1b.

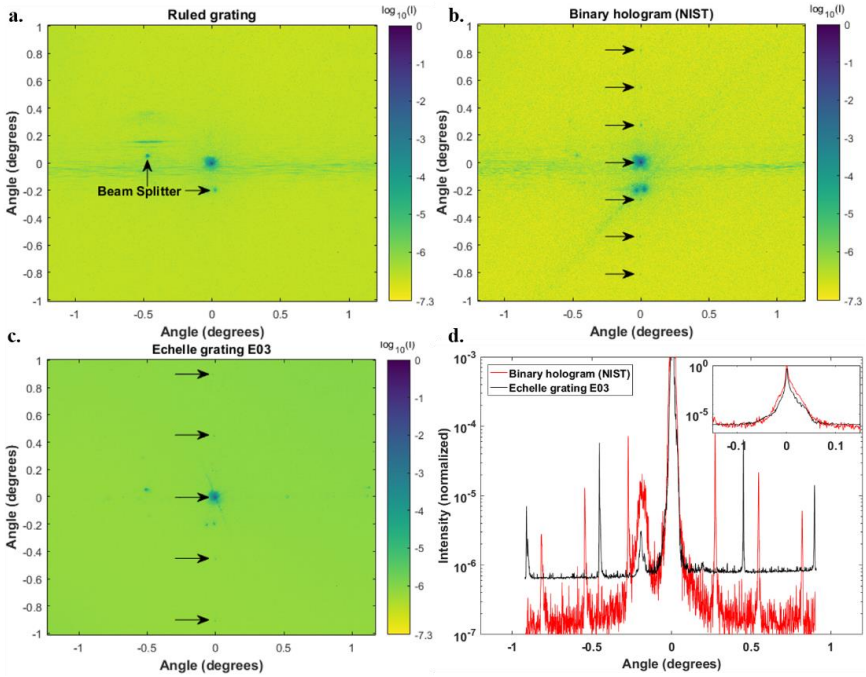


Figure 3. Focal plane images for the 1st diffraction order of the mechanically ruled grating (a), the 1st diffraction order of the binary phase hologram (b), the 26th diffraction order of the echelle grating (c), and sections in the direction of the ordinate through the peak intensity in images b, and c (see graph d). In all images the peak intensity is normalized to one.

3. Measurements and Discussion

Three different gratings were investigated with the focal plane imager. Micrographs of the gratings are shown in Fig. 2. The first is a commercial ruled grating with a dimension of 25.4 mm x 25.4 mm and 11° blaze angle (Fig. 2a). The second grating, strictly speaking, not a linear grating but a computer-generated binary phase hologram etched into borosilicate glass, that was made at NIST for measuring the form errors of an elliptic x-ray mirror [3] (Fig. 2b). The hologram has nearly straight zones. The third grating is an immersed silicon echelle grating that was fabricated at the University of Texas at Austin [4]. While the grating is designed to be used in the infrared through the silicon substrate, we used it as a first surface grating at 37° angle of incidence (26th order) in our measurement.

Fig. 3 shows the focal plane HDR images that were obtained for the three gratings with the procedure described in section 2. Several strong reflections from the beam splitter cube faces are visible in all focal plane images in Fig. 3. Images 3a and 3b have a horizontal streak of stray light across the image that could be traced to light from non-axial diffraction orders that was scattered by the mount for the beam splitter and the focusing lens. The most intriguing observation are the small, vertically distributed peaks in Figs. 3b and 3c (arrows). In case of Fig. 3b, we can trace them to an error in the hologram with (66.4 ±0.1) µm period that is caused by the step-and-scan motion of the lithography system that was used to create the hologram. This error was not discernible in interferometric measurements. Fig. 3c suggests that a similar periodic fabrication error is present in the silicon echelle grating.

4. Acknowledgments

We are grateful to Dr. Cynthia Brooks and Ben Kidder of the University of Texas at Austin for the loan of a silicon echelle grating. The phase grating in Fig. 2b was fabricated using resources provided by the Center for Nanoscale Science and Technology (CNST) NanoFab at NIST. Supported in part by NASA APRA grant NNH15AB22I.

5. References

[1] H. B. Babcock and H. W. Babcock, "The ruling of diffraction gratings at the Mount Wilson Observatory", JOSA **41**, 776-786 (1951).
[2] M. Heusinger, M. Banasch, T. Flügel-Paul, and U.-D. Zeitner, "Investigation and optimization of Rowland ghosts in high efficiency spectrometer gratings fabricated by e-beam lithography", Proc. SPIE 9759, 97590A (2016).
[3] Q. Wang, U. Griesmann, J. Soons, and L. Assoufid, "Figure metrology for x-ray focusing mirrors with Fresnel holograms and photon sieves", Optical Fabrication & Testing, Kona, Hawaii 2014, OSA Technical Digest (online) paper OTU4A.5.
[4] C. B. Brooks, B. Kidder, M. Grigas, U. Griesmann, D. W. Wilson, R. E. Muller, D. T. Jaffe, "Process improvements in the production of silicon immersion gratings," Proc. SPIE 9912, 99123Z (2016).

SIMULTANEOUS NEUTRON AND X-RAY IMAGING OF 3D KEROGEN AND FRACTURE STRUCTURE IN SHALES

Wei-Shan Chiang[†,‡,⊥], Jacob M. LaManna[¶], Daniel S. Hussey[¶], David L. Jacobson[¶], Yun Liu[‡,⊥,#], Jilin Zhang[†], Daniel T. Georgi[†], and Jin-Hong Chen[†]

[†]Reservoir Engineering Technology, Aramco Research Center – Houston

[‡] Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland

[⊥] Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware, USA

[¶]Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland

[#]Department of Physics and Astronomy, University of Delaware, Newark, Delaware, USA

This paper was prepared for presentation at the SPWLA 58th Annual Logging Symposium held in Oklahoma City, Oklahoma, USA, June 17-21, 2017.

ABSTRACT

Hydrocarbon production from shales using horizontal drilling and hydraulic fracturing has been the key development in the US energy industry in the past decade and has now become more important globally. Nevertheless, many fundamental problems related to the storage and flow of light hydrocarbons in shales are still unknown. It has been reported that the hydrocarbons in the shale rocks are predominantly stored within the kerogen pores with characteristic length scale between 1 nm to 100 nm. In addition, the 3D connectivity of these kerogen pores and fractures from the micrometer to centimeter scale form the flow path for light hydrocarbons. Therefore, to better model the gas-in-place and permeability in shales, it is necessary to quantify the structural distribution of organic and inorganic components and fractures over a large breadth of length scales.

Simultaneous neutron and X-ray tomography offers a core-scale non-destructive method that can distinguish the organic matter, inorganic minerals, and open and healed fractures in 2.5 cm diameter shales with resolution of about 30 μm and field of view of about 3 cm. In the reconstructed neutron volume, the hydrogen-rich areas, i.e. organic matter, are brighter because hydrogen has a larger attenuation coefficient and attenuates neutron intensity more significantly. For the X-ray volume, the attenuation coefficient of an element is related to its atomic number Z and the brighter areas indicate the region containing more high-Z elements such as minerals. Open fractures do not attenuate either neutrons or X-rays and therefore look dark in both reconstructed neutron and X-ray volumes.

In this study, two shale samples from different locations were investigated using simultaneous neutron and X-ray tomography for the first time. We were able to construct 3D images of shales and isolate 3D maps of organic matter and high-Z minerals. The distribution of kerogen and fractures can be used in the modeling of hydrocarbon flow in core scale, a 10^9 upscaling from current methods that model the flow based on SEM images.

Q

INTRODUCTION

Natural gas production worldwide from shale resources is projected to grow from 1.2×10^9 m^3/d in 2015 to 4.8×10^9 m^3/d by 2040 (Aloulou and Zaretskaya 2016). Shale gas is predicted to account for 30% of world natural gas production by 2040 (Aloulou and Zaretskaya 2016). Despite the promising future of shale gas as an energy source, fundamental problems related to the storage and transport of hydrocarbons in shales still lack clear understanding. The organic matter in organic rich source rocks may include kerogen, bitumen, and/or heavier immobile hydrocarbons. Kerogen is imbedded within the inorganic matrix of shale rocks and does not dissolve in any solvent (Loucks et al. 2009). Previous studies have suggested that the majority of light hydrocarbons in the shale rocks are stored within the kerogen which has pores with the characteristic length scale between 1 to 100 nm (Ambrose et al. 2013). The total amount and the distribution of kerogens in the shales, therefore, can influence the total gas in place (GIP) of the shales. Moreover, the kerogens and minerals in shales have very

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

different properties such as pore structure (pore size, pore shape, and pore connectivity), surface chemistry, wettability, and mechanical response. These properties are directly related to the interaction of hydrocarbons with pore surfaces. Thus, the structure and three-dimensional (3D) space arrangement of the components within the rocks can significantly influence the flow path of the hydrocarbons in the shales. To better understand the storage and transport properties of hydrocarbons within the shale rocks, it is necessary to identify the distribution and the structure of organic components, especially the kerogens, and inorganic components, i.e. the minerals, and fractures in the rocks.

The development of focused ion beam-scanning electron microscopy (FIB-SEM) enables the visualization of the localized pores, kerogens, and inorganic matrix in small shale samples (Loucks et al. 2009). Moreover, the FIB-SEM images have been used to construct the 3D pore scale representations as the input to the simulation and modeling of hydrocarbon transport in the shales (Jiang et al. 2017; Shabro et al. 2013). However, the structure of the rocks and flow properties obtained from FIB-SEM are at the pore scale. There is also the need to obtain the distribution of organic matter, kerogen and fractures at the core scale.

In the petroleum industry, X-ray computed tomography (CT) is a common tool to characterize the structure of source rocks. X-rays interact with electrons of the materials and, therefore, the attenuation coefficient of X-ray is proportional to the atomic number, Z, of the elements in the materials. High-Z elements such as metals can significantly attenuate X-rays and produce high contrast in the X-ray CT image. This makes X-ray CT a powerful tool to identify the mineral distribution inside the samples. However, organic matter, which is ubiquitous in natural materials, is mainly composed of low-Z elements such as hydrogen and carbon. These materials only have small interaction with X-rays and are fairly transparent to X-rays. What is more, when fractures and pores are present in the samples, it is hard for X-rays to distinguish between these "empty" spaces and the organic matter.

Unlike X-rays, neutrons interact with the atomic nuclei of elements inside the materials. The attenuation coefficient for neutrons has little correlation with Z (Banhart et al. 2010). In particular, hydrogen, which is the most transparent element for X-ray and is very abundant in organic matter, has a very large attenuation coefficient for neutron. The high sensitivity of neutron to organic matter makes neutron imaging a complementary tool to the widely-used X-ray CT for thoroughly and correctly characterizing the structure and components of the heterogeneous shale samples at the core scale. Dual neutron and X-ray tomography has been successfully applied to many other materials such as fuel cells (Banhart et al. 2010; Manke et al. 2007b), batteries (Manke et al. 2007a), and cultural heritage objects (Mannes et al. 2015; Mannes et al. 2014).

In this study, for the first time, "simultaneous" dual X-ray and neutron imaging is used to study shales at the core scale at the National Institute of Standards and Technology (NIST) Center for Neutron Research (NCNR). The distributions of minerals, organic matter and fractures are successfully identified in a Barnett shale and a Middle East shale. The distribution of organic matter is found to have very different configuration and distribution in these two shales.

Q

METHOD

Attenuation of X-ray and neutron by materials. X-ray imaging and neutron tomography are both non-invasive methods and have similar principles. When an incident X-ray or neutron beam with intensity I_0 interacts with a material, the X-ray or neutron intensity is attenuated by different interaction processes. For a homogeneous medium with thickness d, the transmitted intensity of X-rays or neutrons can be described by the Lambert-Beer law:

$$I_t = I_0 \, e^{-\mu \cdot d} \qquad (1)$$

μ is the attenuation coefficient, which is a material property and depends on the incident radiation, i.e. X-ray or neutron, and incident radiation energy. μ can be related to atomic composition of the material by

$$\mu = \sum_i \sigma_{t,i} \, N_i \qquad (2)$$

$\sigma_{t,i}$ is the total attenuation cross section for X-ray or neutron and N_i is the number density of type i

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

atom. μ is very different for X-rays and for neutrons because the X-rays interact with electrons while neutrons interact with nuclei in the materials.

For X-rays, the intensity attenuation is due to Thomson (coherent) scattering, photoelectric absorption, Compton scattering, and pair production. Thus, the total cross section for X-rays for atom i, $\sigma_{i,t}^{X-ray}$, can be written as

$$\sigma_{i,t}^{X-ray} = \sigma_i^{TS} + \sigma_i^{PA} + \sigma_i^{CS} + \sigma_i^{PP} \qquad (3)$$

σ_i^{TS}, σ_i^{PA}, σ_i^{CS}, and σ_i^{PP} are X-ray atomic cross sections for Thomson scattering, photoelectric absorption, Compton scattering, and pair production, respectively. Pair production is ignored because the peak energy of 90 keV of the X-ray tube used in this work is well below the 1.022 MeV threshold for its occurrence.

Neutron has many ways to interact with matter characteristic of different cross sections such as elastic scattering (σ_i^{ES}), inelastic scattering (σ_i^{IES}), radiative capture (σ_i^{RC}), and fission (σ_i^{F}). Thus, the total cross section for neutron for atom i, $\sigma_{i,t}^{Neutron}$, can be written as

$$\sigma_{i,t}^{Neutron} = \sigma_i^{ES} + \sigma_i^{IES} + \sigma_i^{RC} + \sigma_i^{F} + \cdots \quad (4)$$

For heterogeneous materials, the integral form should be used:

$$I_t = I_0\, e^{-\int_l \mu(x)dx} \qquad (5)$$

l denotes the transmission path and x is a 3-dimensional position vector. The transmission of the beam through the sample depends upon the map of the attenuation coefficient $\mu(x)$ and transmission path.

Table 1 lists the attenuation coefficient, μ, for several common elements found in shales for X-ray and neutron at relevant conditions. Attenuation coefficients for neutron and X-ray are calculated by using Neutron Activation and Scattering Calculator provided by NCNR (Kienzle) and X-Ray Form Factor, Attenuation, and Scattering Tables provided by NIST Physical Measurement Laboratory (Seltzer), respectively. Clearly, hydrogen containing materials have large attenuation coefficient for neutrons compared with hydrogen-free elements.

Table 1. The attenuation coefficient of X-ray and neutron for common elements in shales

Component	μ for X-ray 90 keV[a] (cm^2/g)	μ for neutron 1.8 Å[b] (cm^2/g)
Kerogen[c] ($C_{100}H_{161}N_{1.85}S_{0.7}O_{9.2}$)	0.22	6.97
Quartz (SiO_2)	0.48	0.29
Pyrite (FeS_2)	1.66	0.44
Siderite ($FeCO_3$)	1.20	0.67
Calcite ($CaCO_3$)	0.58	0.35
Dolomite ($CaMg(CO_3)_2$)	0.55	0.41
Fluorapatite $Ca_5(PO_4)_3F$	0.71	0.31
Anhydrite ($CaSO_4$)	0.64	0.29
Kaolinite ($Al_2Si_2O_5(OH)_4$)	0.47	2.32
Chlorite ($Mg_5Al_2Si_3O_{10}(OH)_8$)	0.44	2.07
Daphnite ($Fe_5Al_2Si_3O_{10}(OH)_8$)	0.95	2.30
H_2O	0.18	5.65
C_6H_{14}	0.12	5.39

[a]90 keV is the peak X-ray energy for the spectrum generated by the X-ray tube in this measurement. Attenuation coefficient for X-ray is obtained from X-Ray Form Factor, Attenuation, and Scattering Tables (Seltzer).

[b]The neutrons used in this study are thermal neutrons with a Maxwell-Boltzmann distribution centered around 1.8 Å wavelength. Attenuation coefficient for neutron is obtained from Neutron Activation and Scattering Calculator (Kienzle).

[c]Formula obtained from literature (Siskin et al. 1995).

Samples. Two samples were used in this study. The first one, provided by Bureau of Economic Geology (BEG), is a Barnett shale from T.P. Sims #2 well in the Fort Worth Basin. The characterization of this sample can be found in previous study (Louck et all 2009; Zhang et al. 2014). It has Vitrinite Reflectance *Ro* = 1.61 % and total organic carbon (TOC) = 3.64 %. A

Q

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

second sample is an organic-rich carbonate source rock from the Middle East and shall be named Middle East shale. The samples are 2.5 cm in diameter and approximately 2 cm in length.

Experimental set-up. Experiments were conducted at NCNR on the BT2 neutron imaging facility. This instrument offers a high thermal neutron flux for radiography and tomography. A 90 keV microfocus X-ray generator is located perpendicular to the neutron beam at the sample location to allow simultaneous neutron and X-ray tomography. Full details of the instrument can be found in the following manuscript (LaManna et al. 2017). Detector resolution was set to 30 µm for both neutron and X-ray detectors to keep the sample in the field-of-view at all projection angles. Each scan took approximately 18 hours to complete. Volumes were reconstructed from the projection images using the commercial CT software package Octopus[1] and visualized with the open-source Drishti Volume Exploration Tool (Limaye 2012).

RESULTS

The two shales were studied using the simultaneous X-ray and neutron tomography to extract the 3D structure and spatial arrangement of their components such as minerals, organic matter, and fractures in the rocks.

Barnett shale. Figure 1 shows the simultaneous X-ray and neutron CT reconstruction slices for the Barnett shale. The brighter parts of the image indicate regions with larger attenuation, while the darker parts correspond to regions with less attenuation. For neutrons, the hydrogen-rich areas are brighter because hydrogen has larger cross section and attenuates neutron intensity more significantly. For X-rays, the cross section of an element is proportional to its atomic number Z and the brighter areas indicate the regions containing more high-Z elements such as heavy minerals.

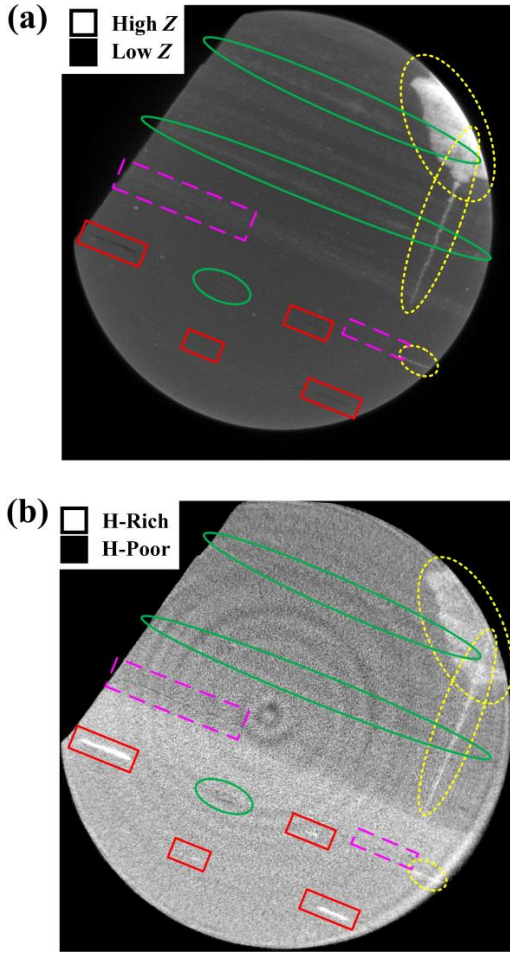


Fig. 1 Upper slices of (a) X-ray and (b) neutron reconstructed volumes of the Barnett shale sample. The short-dashed yellow ovals and solid green ovals outline the mineral components of pyrite and apatite, respectively. Solid red rectangles outline the organic-rich regions likely to be kerogens, the long-dashed magenta rectangles outline the open fractures.

Layered patterns are apparent in the upper part of the sample in both X-ray and neutron tomography slices (Figure 1). The bright layers in the X-ray slice indicate that these layers have high atomic number and, therefore, are composed of minerals (see the layers outlined by the solid green ovals). These bright layers in X-ray slice correspond to

[1] Certain trade names and company products are mentioned in the text or identified in an illustration in order to adequately specify the experimental procedure and equipment used. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

the dark layers in neutron slice. There is also a bright pocket connected with a bright sharp line on the upper right of the slice for both neutron and X-ray images (outlined by short-dashed yellow ovals). This indicates that element in this region has high atomic number and at the same time can strongly attenuate neutron intensity. This area can possibly be occupied by minerals with high mass density iron-rich components which exist in common shale minerals. It should be noticed that the bright layers and bright pocket shown in X-ray slice are composed of different minerals. These two minerals can be easily distinguished in the neutron slice where the layers look dark while the pocket looks bright. The large bright pocket in both X-ray and neutron images is pyrite. The stripes bright in X-ray but dark in neutron are Fluorapatite (see Table 1 for attenuation coefficients and Appendix for XRD result).

Neutron image shows sporadic bright stripes which correspond to the dark stripes in the X-ray image as outlined by the solid red rectangles in Figure 1. The stripes are regions with concentrated hydrogen which attenuate neutrons but are transparent to X-rays. The hydrogen-rich areas are quite likely to be organic kerogen. There are also regions that look dark in both neutron and X-ray images in Figure 1 (outlined by the long-dashed magenta rectangles). Because the "element" is transparent to and therefore has no interaction with both neutron and X-ray, it is reasonable to identify the lines within the long-dashed magenta rectangles in Figure 1 as fractures. If X-ray is the only imaging source (Figure 1(a)), it is hard to distinguish between the organic matter (solid red rectangles) and the fractures (long-dashed magenta rectangles). With the aid of neutron imaging, the two different components can be easily recognized.

The lower part of the neutron slice is homogenously brighter than upper part, indicating that the hydrogen content is higher in the lower part rather than in the upper part of the sample. This is consistent with the fact that the upper part is composed of the layered and pocket inorganic matter containing minerals.

Figure 2(a) and 2(b) are the 3D X-ray and neutron volume renderings of the T.P. Sims #2 well sample, respectively. The colored regions show the highly attenuated areas. The rock sample can be divided into two parts: one with the mineral-rich region containing layers of Fluorapatite and a pocket connecting to a healed fracture filled with

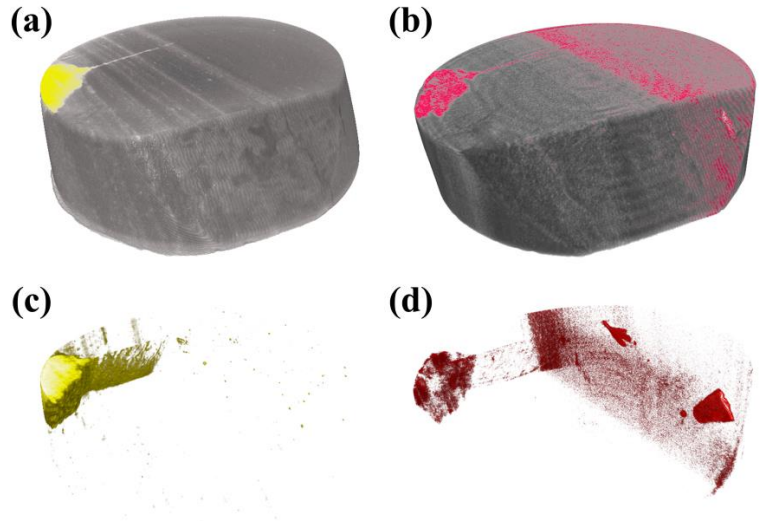


Fig. 2 3D volume renderings of the Barnett shale sample for (a) X-ray and (b) neutron tomography. Distributions of (c) high-Z components, i.e. the minerals, showing a pyrite pocket and a healed fracture; and (d) hydrogen-rich components, likely to be kerogens, are identified by isolating the appropriate peak in the grayscale histograms.

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

pyrite; the other with a relatively homogeneous hydrogen-rich region similar to that seen in the slices shown in Figure 1. By setting a high gray-scale display threshold, the high-attenuation regions, i.e. the yellow and red areas are selected from the X-ray volume and neutron volume effectively identifying the densest inorganic or mineralized (Figure 2(c)) and main organic rich regions (Figure 2(d)), respectively. Figure 2(c) shows the pyrite pocket and the pyrite healed fracture that cuts through the layered structure at an oblique angle. Figure 2(d) highlights several areas with highly concentrated hydrogen content that corresponds to voids in the X-ray in addition to the overall dispersed organic content in the one half of the sample. It should be noted that even though the pocket and healed fracture in mineral-rich side of the rock highly attenuate neutron, they are not organic matter but are pyrite (see Figure 2(c)). The 3D structure of organic matter allows us to map the connection of kerogen 'globs' and fractures and, thus, suggests a possible flow path for light hydrocarbons at the core scale in this shale. Therefore, neutron tomography provides unique and significant information for modeling flow in shales in core scale.

Middle East shale. Figure 3 shows the simultaneous X-ray and neutron tomography slices for the Middle East shale. As described above, the regions outlined by solid green ovals are bright in the X-ray image but dark in the neutron image and, therefore, are occupied by minerals. There is a region outlined by short-dashed yellow oval that is bright in both X-ray and neutron images and is also mineral-rich. Based on the same reasons described for Barnett shale sample, these minerals may be anhydrite and pyrite for areas outlined by solid green ovals and short-dashed yellow oval, respectively. The regions outlined by solid red rectangles are dark on the X-ray image but bright in neutron image and likely are organic matter. There is a stripe which is dark in both neutron and X-ray images and is identified as fracture (outlined by long-dashed magenta rectangles).

Figure 4(a) and 4(b) are the 3D reconstructed images of the Middle East shale for X-ray and neutron modes, respectively. The colored regions are highly attenuated areas. By only displaying the high-attenuation regions, the distributions of inorganic and organic regions are plotted in Figure 4(c) and 4(d), respectively. Surprisingly, the organic matter has a clear layered structure with two layers at the top and in the center of the sample. This is very different from the structure of the organic matter found in the Barnett shale. Moreover, the inorganics are also clearly layered and form a continuous inorganic-rich layered shale (a.k.a. mud rock) (Figure 4(c)) and organic (Figure 4(d)) layers in the top and center of the sample.

The layered structures of organic matter have two important implications: first, it may give rise to fast path for light hydrocarbon movement. Second, it produces strong contrasts in mechanical properties which in turn may be weak surfaces along which the sample may break. The layer at the top of the sample, for example, is a broken face during plugging.

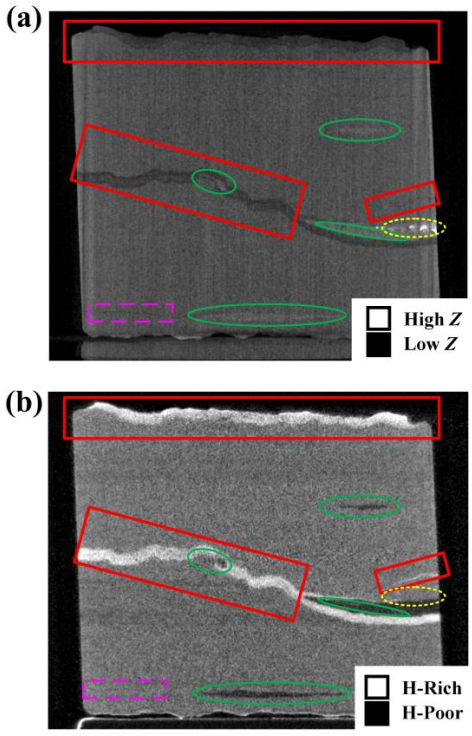


Fig. 3 Central slices of (a) X-ray and (b) neutron reconstructed volumes of the Middle East shale sample. The short-dashed yellow ovals and solid green ovals outline the mineral components of pyrite and probable anhydrite, respectively. Solid red rectangles outline the organic-rich regions

Q

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

likely to be kerogens; the long-dashed magenta rectangles outline the fractures.

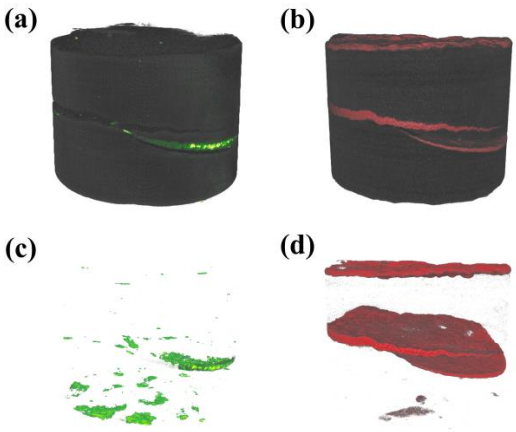


Fig. 4 3D volume renderings of the Middle East shale sample for (a) X-ray and (b) neutron tomography. Distributions of (c) high-Z components, i.e. the minerals, and (d) hydrogen-rich components, likely to be kerogens, are identified from the highly attenuated (colorized) regions in 3D X-ray and neutron tomography reconstructions, respectively.

DISCUSSIONS

The distributions of different components within shales are crucial for storage and transportation of hydrocarbons in the shales. The detailed structural description on the core scale of two shale samples demonstrates that simultaneous X-ray and neutron tomography is a powerful tool to quantify the distributions of organic and inorganic matter. With more information provided by the combined X-ray and neutron tomography, different minerals such as pyrite, calcite, and anhydrite can be distinguished. Moreover, since the hydrogen-rich organic matter and fractures are both very transparent to X-rays, it is challenging for the commonly used X-ray CT alone to unambiguously determine the organic matter distributions at the core scale. In contrast, neutrons are very sensitive to hydrogens and are useful to characterize the organic-rich shales. By combining X-ray and neutron tomography, more comprehensive information of material composition and distributions in shales are obtained.

It is worth mentioning that the component distributions found in this study are at the core scale, rather than the nanometer to micrometer pore scale. Current simulation and modeling of hydrocarbons are mostly based on the rock structure in pore scale constructed from SEM images (Chen et al. 2015; Jiang et al. 2017; Saraji and Piri 2015; Shabro et al. 2013). There is about a 10^9 order of magnitude upscaling in length scale from the pore to core scale. It will be a significant impact to study the influence of the distributions of different components within the rocks by simulating and modeling the hydrocarbon flow on core scale.

Since the wettability and surface properties are very different for the minerals and organic matter (Saraji et al. 2017), the structure and arrangement in space of these components can directly affect the storage and transportation of the hydrocarbons within the shale rocks (Chiang et al. 2016a; Chiang et al. 2016b). The two shale samples from two different formations exhibit very different organic matter structures. The laminar organic matter (kerogens) in the Middle East shale can dramatically change the current views of hydrocarbon flow and the mechanical properties in the shales.

Q

CONCLUSIONS

The simultaneous X-ray and neutron tomography is necessary to more accurately characterize the distributions of different components such as minerals, organic matter, and fractures in the heterogeneous materials. In this study, the 3D characterizations of minerals and organic matter at the core scale are successfully demonstrated for the first time in shale samples from different locations. These results can be used in future simulation and modeling of hydrocarbon flow in core scale. In particular, the layered organic structure found in one of the shale samples can significantly change the current explanations of hydrocarbon flow in the shales.

ACKNOWLEDGEMENTS

The authors would like to thank Aramco for allowing the publication of this work. The technical work has benefited from discussions with colleagues at Aramco Research Center-Houston: Younane Abousleiman, Hui-Hai Liu.

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

REFERENCES

Aloulou, F., Zaretskaya, V.: Shale gas production drives world natural gas production growth - Today in Energy - U.S. Energy Information Administration (EIA), https://www.eia.gov/todayinenergy/detail.php?id=27512.

Ambrose, R.J., Hartman, R.C., Diaz Campos, M., Akkutlu, I.Y., Sondergeld, C.: New Pore-scale Considerations for Shale Gas in Place Calculations. In: SPE Unconventional Gas Conference. Society of Petroleum Engineers (2013).

Banhart, J., Borbély, A., Dzieciol, K., Garcia-Moreno, F., Manke, I., Kardjilov, N., Kaysser-Pyzalla, A.R., Strobl, M., Treimer, W.: X-ray and neutron imaging – Complementary techniques for materials science and engineering. Int. J. Mater. Res. 101, 1069–1079 (2010).

Chen, L., Zhang, L., Kang, Q., Viswanathan, H.S., Yao, J., Tao, W.: Nanoscale simulation of shale transport properties using the lattice Boltzmann method: permeability and diffusivity. Sci. Rep. 5, 8089 (2015).

Chiang, W.-S., Fratini, E., Baglioni, P., Chen, J.-H., Liu, Y.: Pore Size Effect on Methane Adsorption in Mesoporous Silica Materials Studied by Small-Angle Neutron Scattering. Langmuir. 32, 8849–8857 (2016)(a).

Chiang, W.-S., Fratini, E., Baglioni, P., Georgi, D., Chen, J., Liu, Y.: Methane Adsorption in Model Mesoporous Material, SBA-15, Studied by Small-Angle Neutron Scattering. J. Phys. Chem. C. 120, 4354–4363 (2016)(b).

Jiang, W., Lin, M., Yi, Z., Li, H., Wu, S.: Parameter Determination Using 3D FIB-SEM Images for Development of Effective Model of Shale Gas Flow in Nanoscale Pore Clusters. Transp. Porous Media. 117, 5–25 (2017).

Kienzle, P.: Neutron Activation Calculator, https://www.ncnr.nist.gov/resources/activation/.

LaManna, J.M., Hussey, D.S., Baltic, E., Jacobson, D.L.: Neutron and X-ray Tomography (NeXT) system for simultaneous, dual modality tomography. Submitted to Review of Scientific Instruments (2017).

Limaye, A.: Drishti: a volume exploration and presentation tool. In: Stock, S.R. (ed.) Developments in X-Ray Tomography VIII. p. 85060X. International Society for Optics and Photonics (2012).

Loucks, R.G., Reed, R.M., Ruppel, S.C., Jarvie, D.M.: Morphology, Genesis, and Distribution of Nanometer-Scale Pores in Siliceous Mudstones of the Mississippian Barnett Shale. J. Sediment. Res. 79, 848–861 (2009).

Manke, I., Banhart, J., Haibel, A., Rack, A., Zabler, S., Kardjilov, N., Hilger, A., Melzer, A., Riesemeier, H.: *In situ* investigation of the discharge of alkaline Zn–MnO2 batteries with synchrotron x-ray and neutron tomographies. Appl. Phys. Lett. 90, 214102 (2007)(a).

Manke, I., Hartnig, C., Grünerbel, M., Lehnert, W., Kardjilov, N., Haibel, A., Hilger, A., Banhart, J., Riesemeier, H.: Investigation of water evolution and transport in fuel cells with high resolution synchrotron x-ray radiography. Appl. Phys. Lett. 90, 174105 (2007)(b).

Mannes, D., Benoît, C., Heinzelmann, D., Lehmann, E.: Beyond the Visible: Combined Neutron and X-ray Imaging of an Altar Stone from the Former Augustinian Church in Fribourg, Switzerland. Archaeometry. 56, 717–727 (2014).

Mannes, D., Schmid, F., Frey, J., Schmidt-Ott, K., Lehmann, E.: Combined Neutron and X-ray Imaging for Non-invasive Investigations of Cultural Heritage Objects. Phys. Procedia. 69, 653–660 (2015).

Saraji, S., Piri, M.: The representative sample size in shale oil rocks and nano-scale characterization of transport properties. Int. J. Coal Geol. 146, 42–54 (2015).

Saraji, S., Piri, M., Akbarabadi, M., Georgi, D., Delshad, M.: Nano-scale Experimental Investigation of In-situ Wettability and Spontaneous Imbibition in Ultra-tight Reservoir Rocks. Submitted to Water Resources Research (2017).

Seltzer, S.M.: NIST X-Ray Form Factor, Atten. Scatt. Tables Form Page, http://physics.nist.gov/PhysRefData/FFast/html/form.html.

Shabro, V., Kelly, S., Torres-Verdín, C., Sepehrnoori, K.: Pore-Scale Modeling of Electrical Resistivity and Permeability in FIB-SEM Images of Hydrocarbon-

Q

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

Bearing Shale. In: SPWLA 54th Annual Logging Symposium, 22-26 June, New Orleans, Louisiana. Society of Petrophysicists and Well-Log Analysts (2013).

Siskin, M., Scouten, C.G., Rose, K.D., Aczel, T., Colgrove, S.G., Pabst, R.E.: Detailed Structural Characterization of the Organic Material in Rundle Ramsay Crossing and Green River Oil Shales. In: Composition, Geochemistry and Conversion of Oil Shales. pp. 143–158. Springer Netherlands, Dordrecht (1995).

Zhang, T., Yang, R., Milliken, K.L., Ruppel, S.C., Pottorf, R.J., Sun, X.: Chemical and isotopic composition of gases released by crush methods from organic rich mudrocks. Org. Geochem. 73, 16–28 (2014).

APPENDIX

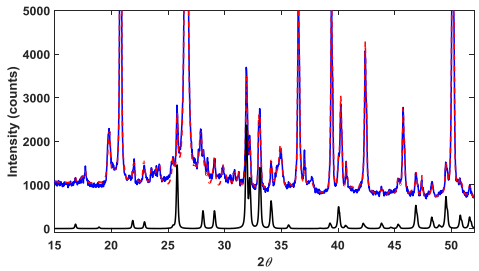


Fig. A1 XRD of the Barnett shale sample. The blue spectrum is the diffractogram from a Bruker D8 Advance Eco (at 40kV 25 mA with a 0.6mm divergent slit; two 2.5° soller slits; and a LynxEye Detector with all 192 channels used). The red trace is the simulated pattern using Bruker's Topas with the reported minerals; the black trace shows the simulated pattern for the mineral Fluorapatite. The amounts of the minerals are validated using another Rietveld software Autoquan/BGMN.

Q

Chen, Jin-hong; Chiang, Wei-Shan NMN; Georgi, Daniel T.; Hussey, Daniel; Jacobson, David; LaManna, Jacob; Liu, Yun; Zhang, Jilin.
"Simultaneous Neutron and X-Ray Imaging of 3D Kerogen and Fracture Structure in Shales."
Paper presented at SPWLA 58th Annual Logging Symposium, Oklahoma City, OK, United States. June 17, 2017 - June 21, 2017.

Deep Silicon Etching for X-Ray Diffraction Devices Fabrication

Houxun Miao[1], Mona Mirzaeimoghri[1, 2], Lei Chen[3], and Han Wen[1]
[1] National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA
[2] Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA
[3] Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
E-mail: houxun.miao@nih.gov

***Abstract*—We report deep reactive ion etching of silicon gratings via cryogenic and Bosch processes. An aspect ratio of > 50 is achieved for 400 nm period gratings with both processes.**

Keywords—cryogenic silicon etching, Bosch process, x-ray gratings

I. Introduction

X-ray modalities account for the majority of medical imaging procedures worldwide. The partial absorption of x-ray that passing through the object remains the primary contrast mechanism since the invention of x-ray in 1895. The absorption contrast for soft tissues is very low in the hard x-ray regime (20 keV to 100 keV) used in medical imaging. X-ray phase shift introduced by low absorption materials is in principle 3 orders of magnitude higher than the linear attenuation of the amplitude.

We recently developed an x-ray polychromatic far field interferometer [1] using home-fabricated x-ray diffraction gratings of submicron periods. With the phase and decoherence contrasts, we demonstrated a sensitivity improvement of more than an order of magnitude compared to a commercial digital mammographic scanner. Small period gratings promotes high sensitivity and/or compact setup. However, the fabrication of submicron and deep submicron period hard x-ray gratings is still a challenging task, especially for high photon energy x-rays, owing to the fact that the refractive index difference is extremely small between different materials. The small refractive index contrast requires high aspect ratio gratings. For examples, for Au/Si gratings, a π phase shift at 30 keV requires 7.0 μm depth, while for Si/air gratings, the required depth is 38.5 μm.

We fabricated 200 nm and 400 nm periods x-ray gratings by deep reactive etching of Si gratings, coating Pt on the entire grating surface as seed layer with Al_2O_3 in between as adhesion layer, and filling the trenches with Au via conformal electro-deposition [2]. Deep silicon etching is a crucial step in the fabrication process. Beyond x-ray devices, deep Si etching is an important step in many fabrication processes with broad applications in microelectromechanical systems (MEMS), deep trench capacitors, and through-silicon-via packaging, etc. Here we report our work on deep reactive ion etching of silicon gratings using cryogenic and "Bosch" processes. For 200 nm period gratings, an aspect ratio of 44 was achieved using cryogenic process. For 400 nm period gratings, an aspect ratio of > 50 was achieved with both cryogenic and Bosch processes.

II. Sample Preparation

Cr-on-polymer masks were patterned for cryogenic process. Grating patterns of 200 nm or 400 nm periods were transferred from master templates to resist spin coated on silicon wafers via nanoimprint lithography. The master templates were patterned via interference lithography and reactive ion etch of silicon. Cr layers of 10 nm each were deposited via electron beam evaporation at a 30° incident angle to the wafer surface from each side of the grating lines. The residual resist layer was removed via O_2 plasma.

SiO_2 masks were patterned for Bosch process. A layer of 300 nm SiO_2 was grown on a silicon wafer in a wet oxidation furnace and 30 nm Cr was coated via electron beam evaporation. Then resist was spin coated and grating pattern was transferred via nanoimprint lithography. The residual layer of the resist was removed via O_2 plasma. Unprotected Cr was ion milled through. SiO_2 was etched through using a C_4F_8-O_2 recipe and residual Cr mask was cleaned using O_2 plasma.

III. Cryogenic Process

Cryogenic process utilizes a plasma of combined SF_6 and O_2 at low temperature (typically -130 °C to -100 °C) to simultaneously passivate the sidewall and etch the bottom of the silicon [3]. The continuous etching process results in vertical sidewalls without observable roughness.

Two different recipes were created for 200 nm period and 400 nm period grating etching. For 200 nm period gratings, a combination of low inductively coupled plasma (ICP) power and low pressure minimized the undercut underneath the mask and allowed deep silicon etch. With a recipe of 700 W ICP power, 30 W radio-frequency (RF) power, 0.8 Pa pressure, 6.0 x 10^{-7} m^3/s SF_6 and 1.3 x 10^{-7} m^3/s O_2 flow rates, an etching depth of 4.4 μm was achieved in 9 min, corresponding to an aspect ratio of 44. Fig. 1 shows the cross-section scanning electron microscope (SEM) of the etched grating. The small holes in the etching mask were created as the Cr on top of the resist was punched through at the end of the 9 min etch.

Chen, Lei; Miao, Houxun; Mirzaeimoghri, Mona; Wen, Han.
"Deep Silicon Etching for X-Ray Diffraction Devices Fabrication."
Paper presented at 2017 International Conference on Optical MEMS and Nanophotonics, Santa Fe, NM, United States. August 13, 2017 - August 17, 2017.

For 400 nm period gratings, a combination of slightly higher ICP power and pressure promotes faster and deeper etching. With a recipe of 1000 W ICP power, 10 W RF power, 1.1 Pa pressure, 8.7 x 10^{-7} m^3/s SF_6 and 1.3 x 10^{-7} m^3/s O_2 flow rates, an etching depth of 10.6 μm was achieved in 10 min, corresponding to an aspect ratio of 53. Fig. 2 shows the cross section SEM of the etched grating.

For both 200 nm period and 400 nm period gratings, the achievable aspect ratio was ultimately limited by the undercut underneath the mask material. We investigated the effect of different mask materials on the undercut and concluded that Cr-on-polymer mask provides the best trade-off between etching selectivity and undercut [4].

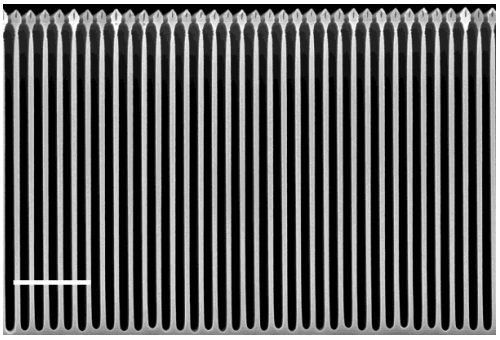


Fig. 1. Cross section SEM image of a 200 nm period grating etched to 4.4 μm via cryogenic process. Scale bar: 1 μm.

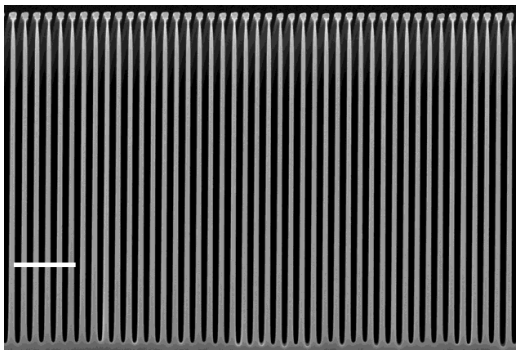


Fig. 2. Cross section SEM image of a 400 nm period grating etched to 10.6 μm via cryogenic process. Scale bar: 2 μm.

IV. Bosch Process

Bosch process alternates SF_6 plasma etching and C_4F_8 plasma passivation steps to obtain directional etching [3]. The switching between etching and deposition steps results in intrinsic scallops. By carefully tuning the etching parameters, the scallops can be reduced. We developed a recipe for 400 nm period grating etching as shown in Table 1. Fig. 3 shows the cross-section SEM image of a grating etched to 11 μm, corresponding to an aspect ratio of 55.

Although, the smoothness and the etching profile are worse than cryogenic process, the undercut underneath the mask is much smaller, which indicates the potential of Bosch process for deeper etching.

TABLE I. Bosch process recipe

Parameters	Steps		
	Deposition	***Etch1***	***Etch2***
ICP Power (W)	2500	2500	2500
RF Power (W)	0	50	40
Pressure (Pa)	6.0	3.3	6.7
C_4F_8 (m^3/s)	6.0 x 10^{-6}	0	0
SF_6 (m^3/s)	0	5.0 x 10^{-6}	6.7 x 10^{-6}
Time (s)	1.6	1	1

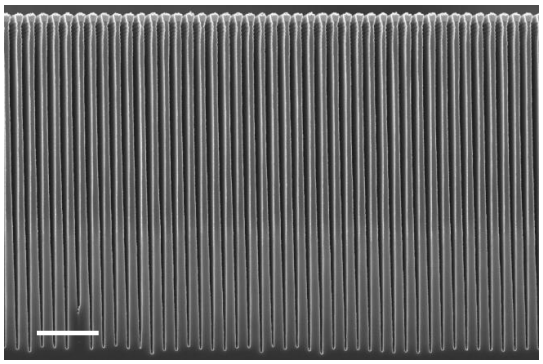


Fig. 3. Cross section SEM image of a 400 nm period grating etched to 11 μm via Bosch process. Scale bar: 2 μm.

V. Conclusion

We demonstrated deep reactive ion etching of submicron silicon gratings via cryogenic and "Bosch" processes. An aspect ratio of 44 was achieved for 200 nm period gratings with cryogenic process. An aspect ratio of > 50 was achieved with both cryogenic and Bosch processes.

Acknowledgment

We acknowledge staffs from the Center for Nanoscale Science and Technology at the National Institute of Standards and Technology for their assistant in fabrication.

References

[1] H. Miao, et al., "A universal moiré effect and application in X-ray phase-contrast imaging," Nature Phys., vol. 12, pp. 830-834, September 2016.

[2] H. Miao, A. A. Gomella, N. Chedid, L. Chen and H. Wen, "Fabrication of 200 nm period hard x-ray phase gratings," Nano Lett., vol. 14, pp. 3453-3458, May, 2014.

[3] B. Wu, A. Kumar and S. Pamarthy, "High aspect ratio silicon etch: a review, " J. Appl. Phys., vol. 108, pp. 051101, September 2010.

[4] H. Miao, L. Chen, M. Mirzaeimoghri, R. Kasica and H. Wen, "Cryogenic etching of high aspect ratio 400-nm pitch silicon gratings, " J MEMS, vol. 25, pp. 963-967, October 2016.

Chen, Lei; Miao, Houxun; Mirzaeimoghri, Mona; Wen, Han.
"Deep Silicon Etching for X-Ray Diffraction Devices Fabrication."
Paper presented at 2017 International Conference on Optical MEMS and Nanophotonics, Santa Fe, NM, United States. August 13, 2017 - August 17, 2017.

High Quality Oxide Films Deposited at Room Temperature by Ion Beam Sputtering

Gerard E. Henein [a*], Juraj Topolancik [b]

a. Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD 20899,
b. Roche Sequencing Solutions, 4155 Hopyard Dr., Suite 200, Pleasanton, CA 94588,
*email: ghenein@nist.gov

Abstract

We have deposited dense and pinhole-free thin films of SiO_2, Al_2O_3 and ITO at room temperature via ion beam sputtering. The SiO_2 films were found to be of similar quality as thermal oxide with a resistivity greater than 10^{13} Ω·m and breakdown field in excess of 700 MV/m. The Al_2O_3 films were part of a Pt- Al_2O_3-Pt vertical tunnel junction and were kept extremely thin, from 2 nm to 4 nm. The current-voltage characteristics of these junctions indicated a breakdown field in excess of 2000 MV/m, roughly twice that achieved by ALD films. This breakdown voltage was found to be independent of junction area, strongly suggesting the absence of pinholes in the film. The ITO films were 50 nm to 100 nm thick. As deposited, they are fully transparent with an electrical resistivity of $5x10^{-6}$ Ω·m.

Introduction

The highest quality oxides such as SiO_2, Al_2O_3 and Indium Tin Oxide (ITO) require high temperature processing either during the growth of the film or annealing post-growth. Thermal SiO_2 is grown at ≈1000 °C [(1)], ALD Al_2O_3 is deposited at ≈300 °C [(2)], and magnetron-sputtered ITO must be annealed above 350 °C in order to turn the film conductive [(3)]. These elevated temperature requirements are not compatible with polymer substrates used for flexible electronics. Our objective was to explore ion beam deposition as an alternative room temperature technique and characterize the film properties.

Film deposition technique

The films of SiO_2, Al_2O_3 and ITO were deposited on silicon wafers at room temperature via ion beam sputtering as depicted in figure 1. The deposition system consists of a 3-grid 14 cm RF ion gun directed at 200 mm

Henein, Gerard; Siebein, Kerry NMN; Topolancik, Juraj.
"High Quality Oxide Films Deposited at Room Temperature by Ion Beam Sputtering."
Paper presented at 2017 MRS Fall Meeting and Exhibits, Boston, MA, United States. November 26, 2017 - December 1, 2017.

targets of SiO_2, Al and ITO. All three processes require a small flow of O_2 to achieve stoichiometry. Typical conditions were: argon flow rate $3.3x10^{-7}$ m^3/s (20 sccm), beam voltage 600 V, beam current 220 mA and acceleration voltage 150 V. The substrate wafers rotated at 10 rpm and were kept at 20 °C. The base vacuum prior to deposition was $2.6x10^{-6}$ Pa ($1.5 x10^{-8}$ Torr).

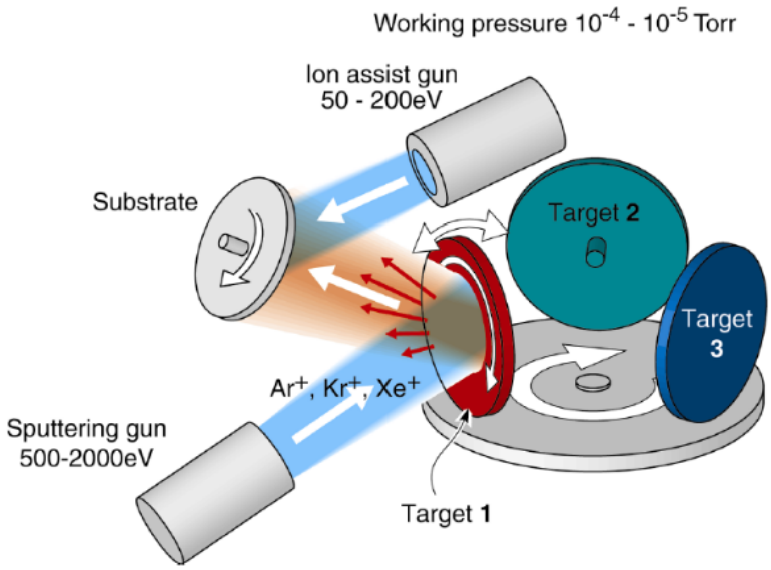


Figure 1. Schematics of the ion beam deposition system

Measurement techniques

The measurements techniques consisted of current-voltage, ellipsometry, chemical etching and X-ray diffraction (XRD). We used a mercury probe directly on the oxide wafers, and a standard probe station on the devices, to measure resistivity, breakdown voltage and detect pinholes. Ellipsometry yielded the index of refraction and extinction coefficient. Chemical etching revealed general oxide quality and checked for the presence of pinholes. XRD was used to assess the degree of crystallinity of the deposited films.

Results

SiO_2 – 100 nm films

Figure 2 shows the index of refraction and extinction coefficients. The index is somewhat higher (by ~0.04) than that of thermal oxide and the extinction coefficient (absorption) nearly zero.

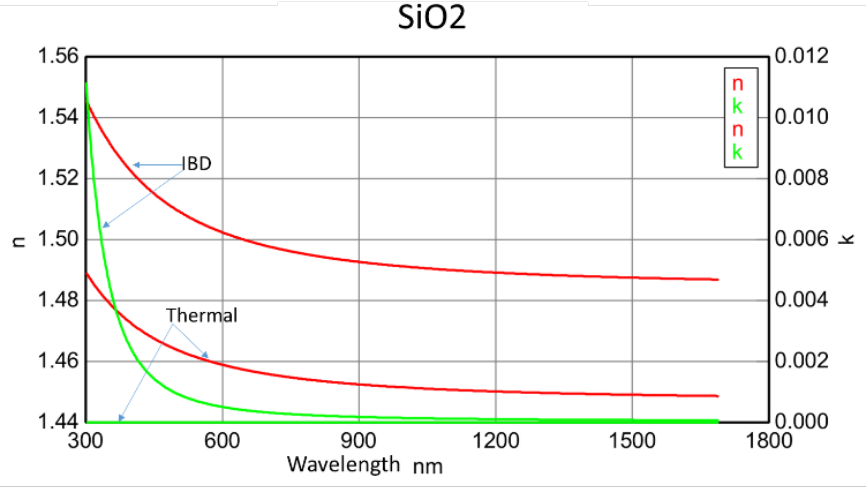


Figure 2. Index of refraction and extinction coefficient of IBD oxide and thermal oxide

We measured the density of IBD oxide: 2,410 kg/m^3, which lies between that of thermal oxide (2,220 kg/m^3) and that of quartz (2,650 kg/m^3). The etch rate in buffered oxide etch (6:1 BOE) was 1.6 nm/sec, identical to that of thermal oxide, and therefore strongly suggesting the absence of pinholes in the film. XRD revealed that the films were totally amorphous.

The resistivity versus electric field was measured via mercury probe between zero and 400 MV/m. The IBD oxide (10^{13}-10^{14} ohm.m) proved somewhat inferior to thermal oxide (10^{14}-10^{15} ohm.m).

Figure 3 shows the breakdown field distribution of IBD oxide and thermal oxide, as measured at 65 points on each wafer via mercury probe. The IBD oxide exhibits a very high breakdown field, from 700 to 1,400 MV/m, higher than thermal oxide.

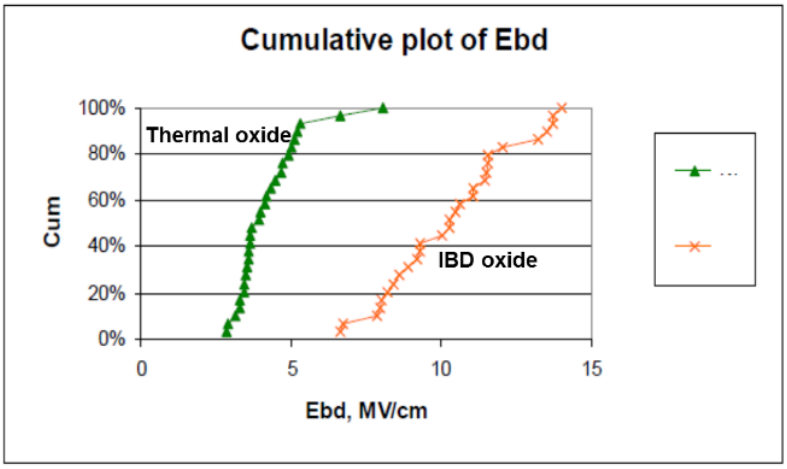


Figure 3. Breakdown field distribution of IBD and thermal oxides

Al_2O_3 – 50 nm films

We measured the index and extinction coefficient in the wavelength range of 280 nm to 1650 nm, and found them identical to those of ALD (Atomic Layer Deposition) oxide. The index is 1.67 at 600 nm while the extinction coefficient is less than 0.01.

The electrical resistivity was measured by a mercury probe, and is in excess of $3x10^{12}$ ohm.m up to fields of 150 MV/m. This value is superior to that of thermal oxide (obtained by oxidation of AlN, ~10^{10} ohmcm). The breakdown voltage was also measured via a mercury probe and found to yield a breakdown field in the range 200-370 MV/m, slightly lower than that of thermal oxide (400-500 MV/m.)

XRD revealed that the films were totally amorphous.

ITO – 100 nm films

The index and extinction coefficient were found nearly similar to those of standard magnetron-sputtered ITO. The electrical resistivity, *as deposited*, was $5x10^{-6}$ ohm.m, similar to that of magnetron-sputtered ITO but *obtained after annealing at 350 °C.* The films were transparent.

In contrast to all other standard deposition techniques, the as-deposited films were crystalline, as shown in figure 4.

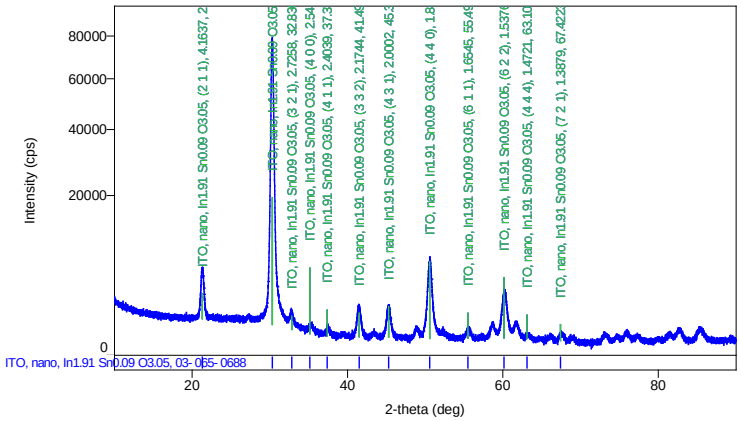


Figure 4. Grazing incidence XRD of ITO film (deposited on glass)

Al_2O_3 – 2-4 nm films

These oxide films were part of a tunnel junction Pt/Al_2O_3/Pt used in a DNA sequencing device. A TEM cross-section of the tunnel junction is shown in figure 5. Of great interest is the sharpness of the interfaces between layers, where interdiffusion can be minimized: first, by using a slow deposition rate for the oxide, and second, by using the Biased Target Deposition technique (BTD) for the titanium and platinum. In BTD, the bias on the metal target can be adjusted to yield a very low adatom energy for the first few monolayers (200-300 eV), thereby preventing interdiffusion, and increasing the bias (to ~800 eV) for the rest of the layer to achieve uniform coverage.

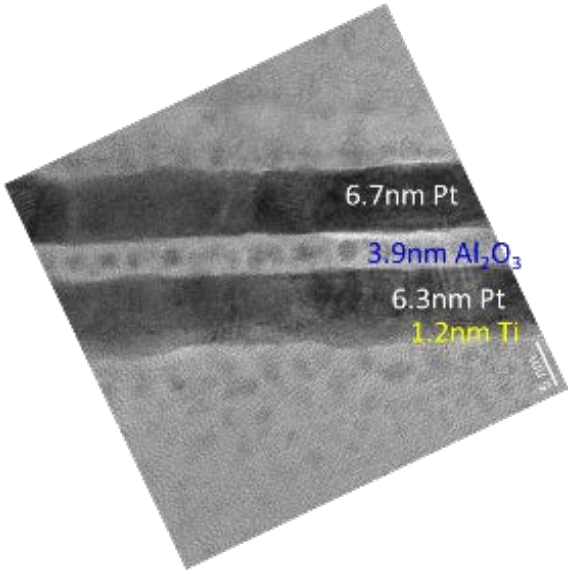


Figure 5. Transmission electron micrograph of tunnel junction cross-section

The oxide thickness was varied from 1 to 8 nm while the junction width was varied from 0.5 to 10 µm. We fabricated junctions with ALD and IBD oxides. We found that the IBD oxide has a superior breakdown field (2,000-3,000 MV/m) to that of ALD oxide (1,300 MV/m). Furthermore, the breakdown field was found to be independent of the junction area, strongly suggesting the absence of pinholes. The data comparing ALD and IBD oxides is shown in figure 6. It is to be noted that, because of the extreme thinness of the junctions, such high breakdown fields arise from quantum tunnelling, and not from material degradation due to defects.

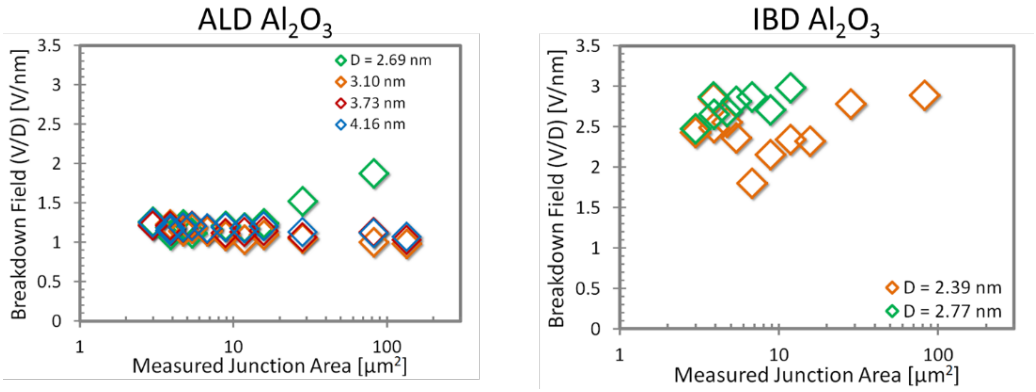


Figure 6. Comparison of breakdown fields for ALD and IBD Al_2O_3

Conclusions

Ion beam deposition is capable of depositing very high-quality oxides at room temperature. The technique allows sub-nanometer control over the film thickness. The biased target version of IBD is able to produce sharp interfaces with minimal interdiffusion.

References

1. E.P. Eernisse, Appl. Phys. Lett. 35(1), 8 (1979)
2. Sun Jin Yun et al., J. Vac. Sci. Tech. 15, 2993 (1997)
3. Yalan Hu et al., Vacuum 75, 183-188 (2004)

Henein, Gerard; Siebein, Kerry NMN; Topolancik, Juraj.
"High Quality Oxide Films Deposited at Room Temperature by Ion Beam Sputtering."
Paper presented at 2017 MRS Fall Meeting and Exhibits, Boston, MA, United States. November 26, 2017 - December 1, 2017.

Energy-Efficient Single-Flux-QuantumBased Neuromorphic Computing

Michael L. Schneider, Christine, A. Donnelly, Stephen E. Russek, Burm Baek, Matthew R. Pufall,
Peter F. Hopkins, William H. Rippard
National Institute of Standards and Technology, Boulder, CO 80305, USA

***Abstract*—Recent experimental work has demonstrated nano-textured magnetic Josephson junctions (MJJs) that exhibit tunable spiking behavior with ultra-low training energies in the attojoule range. MJJ devices integrated with standard single-flux-quantum neural systems form a new class of neuromorphic technologies that have spiking energies between 10^{-18} J and 10^{-21} J, operation frequencies up to 100 GHz, and nanoscale plasticity. Here, we present the design of neural cells utilizing MJJs that form the basic elements in multilayer perception and convolutional networks. We present SPICE models, using experimentally derived Verilog A models for MJJs, to assess the performance of these cells in simple neural network structures. Modeling results indicate that the tunable Josephson critical current I_C can function as a weight in a neural network. Using SPICE we model a fully connected two layer network with 9 inputs and 3 outputs.**

Index Terms—Neuromorphic computing, single flux quantum, magnetic Josephson junctions.

I. Introduction

Many neuromorphic hardware technologies are being explored for their potential to increase the efficiency of computing certain problems and thus facilitate machine learning with greater energy efficiency or with more complexity. Among the technologies being developed, single-flux-quantum-based Josephson junctions are a promising choice for their extremely low energy consumption and intrinsic spiking behavior.

The basic element in single-flux-quantum (SFQ) circuits is the Josephson junction (JJ)[1], which emits a voltage spike with an integrated amplitude equal to the flux quantum ($\varphi_0 = 2.07 \times 10^{-15}\ \mathrm{V \cdot s}$) when the superconducting order parameter undergoes a 2π phase slip[2]. These very low energy pulses have been demonstrated in circuits at frequencies well above 100 GHz[3]. Recently, it has been demonstrated that incorporating a nano-textured magnetic barrier into a Josephson junction results in a JJ whose properties can be dynamically changed via currents and magnetic fields[4]. Here we model neuromorphic circuits based on a combination of traditional SFQ JJs and tunable MJJs using physically realistic parameters.

SFQ circuits are being developed on a large scale for digital supercomputing because of their high speed and extremely low operating energies[5]. This effort can be leveraged in neuromorphic SFQ circuits[6-9]. In this case, we model the neuromorphic circuits with WRSPICE[10].* a version of SPICE that has integrated support for the physics of SFQ circuits. In addition, we used Verilog A to define our own physical model of the nano-textured magnetic JJ. This compact model was compiled and integrated with WRSPICE.

II. Dynamically Tunable Josephson Junction Model

The resistively and capacitively shunted junction (RCSJ) model provides a dynamic equation for the phase evolution of a current-biased Josephson junction[11]. A maximum current of I_C may flow as a supercurrent without any voltage developing across the junction. The relationship between superconducting phase and current is given by the Josephson relation: $I = I_C \sin(\varphi)$, where φ is the phase difference across the junction. In the voltage state, a quasiparticle current contributes to the total current according to $I = G_N(V) \times V$, where G_N is the junction conductivity and V is the junction voltage. The junction voltage V can be related to the phase by the Josephson voltage-phase relationship $d\varphi/dt = 2eV/\hbar$, where e is the electron charge and $\hbar$ is the reduced Planck's constant. . $G_N(V)$ can be strongly voltage-dependent for tunnel junctions or voltage independent for a superconducting-normal-superconducting junction. Here, $G_N(V)$ is assumed to remain constant with voltage for the low resistance junctions in this model. Finally, the displacement current is given by $I = C\, dV/dt$, where C is the junction capacitance, which can again be related to junction phase by the voltage-phase relationship.

The dynamics in the voltage state depend on the Stewart-McCumber parameter[2],

(1) $\beta_C = \frac{2e}{\hbar} I_C R_N^2 C$,

where R_N is the normal state resistance. This parameter is equal to the squared quality factor, Q^2 , of a parallel LRC circuit (here, "L" is the small-signal inductance found by linearizing the supercurrent phase-current relationship about the zero bias phase difference across the junction). The model described in this work uses junctions that are close to critical damping $\beta_C \approx 1$ or overdamped $\beta_C < 1$. For these junctions, a phase evolution in the voltage state followed by a return to the zero-bias phase will cause the junction phase to evolve through one or multiple quantized 2π phase slips. Each of these 2π phase slips can be shown by the Josephson phase-voltage relationship to correspond to a voltage pulse of integrated area $\int V(t)dt = h/2e = \Phi_0$. These phase slips define "single flux quantum pulses"[11].

We modified the RCSJ model that was outlined above to include a dependence of the superconducting critical current on the magnetic order parameter for our MJJ devices[12]. Physically, in the case of the nano-textured magnetic JJs, the

* *This does not imply endorsement by NIST or that the product is the best available for the purpose.*

Baek, Burm; Donnelly, Christine; Hopkins, Peter; Pufall, Matthew; Rippard, William; Russek, Stephen; Schneider, Michael.
"Energy efficient Single Flux Quantum Based Neuromorphic Computing."
Paper presented at 2017 IEEE International Conference on Rebooting Computing (ICRC) , Washington, DC, United States. November 8, 2017 - November 9, 2017.

critical current of the junction can be tuned by changing the order of the magnetic clusters[4]. We model this by defining the magnetic order parameter as

(2) $m = \frac{1}{M}\sum_i \vec{m}_i$

where $\vec{m_i}$ are the moments of each of the magnetic clusters, and M is the total magnetic moment of all clusters in the junction. We define the effect of magnetic order on the junction critical current as:

(3) $I_C(m,T) = \left((1-m)I_{CV} + I_{CM}\right)\left(1 - \left(\frac{T}{T_c}\right)^2\right)$,

where I_{CV} is the variable portion of the critical current, which is influenced by the order of the magnetic particles, I_{CM} is the minimum critical current defined when the barrier has a maximum magnetic order, and T_c is the superconducting transition temperature. The magnetic order parameter varies with the integrated junction voltage according to $dm \propto V(t)dt \mid E_{pulse} > E_m$ between $m = 0$ and $m = 1$, where E_{pulse} is the pulse energy and E_m is the minimum energy required to change the magnetic order. In other words, a positive voltage pulse across the junction, with enough energy, causes the order parameter to increase and the critical current to decrease up to a saturation point. In the circuits presented here, we set the minimum pulse energy equal to 3 aJ, so that once the weights are set, they are not affected by the operational SFQ pulses.

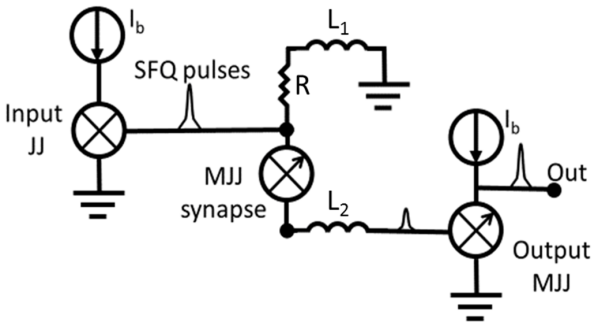


Fig. 1. Circuit diagram of the basic elements connected as 1 input through 1 weight to 1 output. $L_1 = L_2 = 10$ picohenry and $R = 10$ mΩ. The voltage pulse schematics indicate the measured nodes for the input, middle and output neurons.

III. Neuromorphic Unit Cell

Figure 1 shows the schematic of the basic neuromorphic unit cell. The JJ on the left side is performing the function of an input "neuron." If the current bias is on, then the junction will spike. The spiking rate of this junction has been tuned to be roughly 0.35 GHz. The JJ in the center is a "synaptic" MJJ which will weight the transmitted pulse from the input "neuron." We adjust the critical current of these junctions between 1 μA and 100 μA to achieve the desired "synaptic" weighting. This adjustment would be accomplished physically by adjusting the magnetic order in the MJJ, something that can be done dynamically in the system. The JJ on the right side of the circuit acts as the output threshold "neuron". This JJ is also an MJJ but with a higher range of critical currents that we can adjust from 100 μA to 500 μA and is thus analogous to the bias of a neural network. Physically, one can adjust the range of the critical currents that the MJJ has by changing the size of the junction. For the case of 1 – 100 μA currents, devices would be approximately 1 μm in diameter, while MJJs with a critical current range of 5-500 μm can be achieved in MJJs with a diameter of a 3 μm [4].

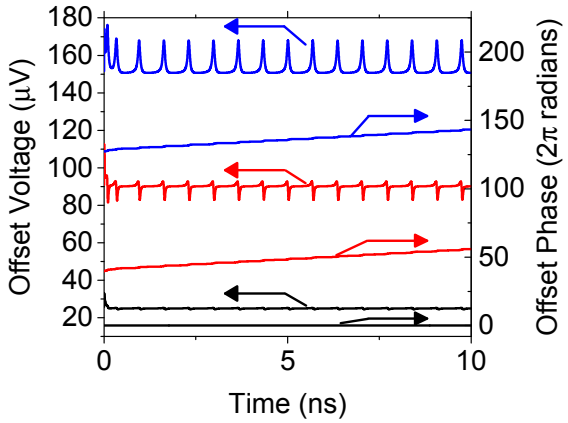


Fig. 2. Behavior of the basic MJJ neural network. The input neuron is on top in blue, the middle MJJ is in red and the output MJJ is in black. All voltages (left axis) are displayed on top of their respective phase evolution (right axis). In this example, the middle MJJ has a critical current of 10 μA corresponding to a low weight, and leading to no spiking on the output.

The circuit operates when the input DC bias is turned on. At this point the JJ on the left will start firing with a frequency of roughly 0.35 GHz. These pulses are then seen as current pulses by the middle synaptic MJJ. This MJJ will fire with a pulse energy ($I_C\varphi_0/2\pi$) that depends on the critical current of this MJJ. The resistor and inductor above the middle MJJ act to stabilize the circuit by introducing a recovery time between spikes, this acts to desensitize the overall circuit from small reflected pulses. The output JJ has a small constant DC bias that is below the spiking threshold. This JJ will undergo a 2π phase slip if the pulse transmitted through the middle JJ is large enough. It should be noted that these bias lines can be shared within a layer as the only energy that should be dissipated is the SFQ energy when the input or output spikes. This has previously been shown to allow for extremely low energy operation in energy efficient rapid single flux quantum circuits[13].

Figures 2 and 3 show the spiking output and phase evolution of the three JJs in the circuit. In Fig. 2 the critical current of the middle MJJ is 10 μA and the output MJJ has an $I_C = 300$ μA. As can be seen in the data, the input and middle JJs are spiking, but the output JJ in this configuration is not spiking. In this example, as the middle junction enters the voltage state, it becomes resistive and more of the input current spike is shunted to the ground above the synaptic JJ. In Fig. 3, the middle MJJ has an $I_C = 80$ μA and the output MJJ has an $I_C = 300$ μA. As can be seen in both the voltage and phase traces, the input and output JJs are firing in this configuration. The weighting JJ in this case

Baek, Burm; Donnelly, Christine; Hopkins, Peter; Pufall, Matthew; Rippard, William; Russek, Stephen; Schneider, Michael.
"Energy efficient Single Flux Quantum Based Neuromorphic Computing."
Paper presented at 2017 IEEE International Conference on Rebooting Computing (ICRC) , Washington, DC, United States. November 8, 2017 - November 9, 2017.

is not undergoing 2π phase slips, but rather is remaining in the superconducting state, and thus passing the initial input pulse through to the output JJ.

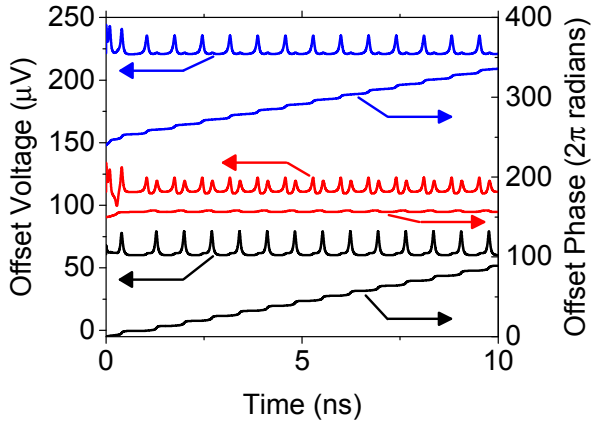


Fig. 3. Behavior of the basic MJJ neural network. The input neuron is on top in blue, the middle MJJ is in red and the output MJJ is in black. All voltages (left axis) are displayed on top of their respective phase evolution (right axis). In this example, the middle MJJ has a critical current of 80 μA corresponding to a high weight, and leading to spiking on the output.

IV. Nine Input Pixel Example

These elements (input JJ, synaptic MJJ, output threshold MJJ) can be used to implement a neural network in hardware. We use an example that was developed as a test for physical memristor circuits[14]. With a 3×3 input pixel array one can define 3 unambiguous letters (z, v, n) and train the neural network to identify these letters and report the answer on one of three outputs. Further, one can add 1 pixel of noise to each of the letters without duplicating patterns that conflate any of the letters. In our example, we use a standard neural network script written in Python to find the weights of a fully connected 9×3 neural network that will be used to solve this example problem[15]. Once the weights are found with the script, we use linear scaling to map these weights to critical current values in the middle layer of the network. In this example, there are only 30 unique input sequences and therefore we use all of the available data for training. As one might expect, we are able to train a standard neural network to reach 100 % identification for this example. In addition, by mapping the weights to critical current values, we are also able to get 100 % correct identification in the SPICE simulation of our MJJ network.

Figure 4 shows the circuit diagram for the JJ neural network. There are 9 input JJs each of which has a DC input bias that is either off (0 μA) or on (500 μA). These inputs are connected to 27 synaptic MJJs which have their critical currents adjusted to form the weights of the middle layer. These are then connected to 3 threshold MJJs forming the output layer.

A SPICE simulation of this circuit was run for 89 ns. In this simulation, the input biases were changed every 2.83 ns to a new test case. After 89 ns, the circuit has gone through all 30 test cases. The circuit correctly identifies the input 100 % of the time. These results are a promising indication that the circuit could be run in real time for image recognition with input images changing as quickly as every 3 ns in this example.

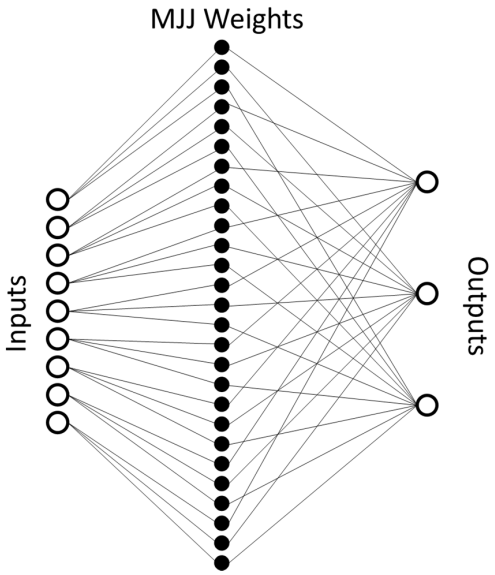


Fig. 4. Schematic of the 9 x 3 MJJ neural network connections.

Figure 5a shows the results of the 89 ns full input image test. The output pulse state for each of the 3 output MJJ neurons is plotted as a function of time. The pattern input is changing every 2.83 ns to the next image input. The output is shown on the graph. The blue output shows spiking on the "z" output neuron, which correctly spikes with the first 10 input cases of "z" or 1 pixel of noise away from "z". The red output shows the spiking on the "v" output neuron, which correctly spikes with the middle 10 input cases of "v" or one pixel away from "v". The black output shows the spiking on the "n" output neuron, which correctly spikes with the last 10 input cases of "n" or 1 pixel of noise away from "n".

V. Discussion

We consider the energy used in this model of spiking JJs by measuring the voltage and current output of each JJ. The value we measure agrees, as expected, with the SFQ pulse energy of $I_C\varphi_0/2\pi$ for each of the JJs. This energy is about 70 zJ for the input JJs, 35 zJ for each of the synaptic JJs and approximately 150 zJ for each of the output JJs. Thus, if we consider only the pulse energies in our network of 39 JJs, this yields an energy dissipation of 2 aJ per classification.

It should be noted however, that this is not the total energy of the system. In a large-scale implementation, there would be an overhead of approximately 1 kW of cooling power consumed per watt dissipated at 4 K. Such a large scale implementation would be capable of processing~ 10^{18} spikes per second in the

Baek, Burm; Donnelly, Christine; Hopkins, Peter; Pufall, Matthew; Rippard, William; Russek, Stephen; Schneider, Michael.
"Energy efficient Single Flux Quantum Based Neuromorphic Computing."
Paper presented at 2017 IEEE International Conference on Rebooting Computing (ICRC) , Washington, DC, United States. November 8, 2017 - November 9, 2017.

ideal case where all energy is dissipated only by spike energy. In our current demonstration there are two other energy dissipation sources that should be considered. The first is that each of the "neuron" layers is DC biased. Since the bias level is not being changed between JJs, this bias current should be able to be shared with a common line on each layer (for example). The other source of power dissipation that should be considered is the resistor in the RL filter that is used to stabilize the circuit. In the current configuration, this resistor is somewhat problematic because it is a source of current dissipation. When the MJJ in the middle layer is firing, this resistor dissipates power that is similar to an additional SFQ pulse energy, which is not problematic, as this would represent a factor of two loss in spike processing. However, there is a small quiescent current that also flows through the resistor in the stabilization element in the present model. The current is a result of the *on* condition being represented using a constant current. By averaging the current and voltage across this resistor for the entire simulation, we observe a dissipation on this resistor of about 5 pW. This amount of dissipation in a stabilizing element will need to be changed if the networks are going to be implemented on a large scale. One potential solution would be to use a switched input where only one pulse is transmitted for the input *on* condition. In this case, the resistor should dissipate at most an additional SFQ pulse energy, which would not be problematic for energy scaling since this energy is already accounted for as the input spike energy.

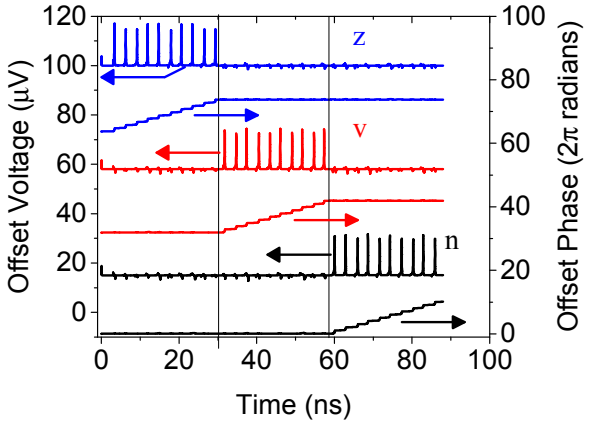


Fig. 5. Voltage spike and phase slip results of the trained 9 x 3 MJJ neural network. 10 spikes in each of the top middle and bottom output MJJs (in that order) corresponding to 100 % classification after training.

VI. Summary

We have demonstrated a 9 input pixel simulation of a neural network based on a combination of traditional JJs and tunable MJJs. We have shown that, with physically realistic parameters, we can solve a simple image recognition problem including recognition in the presence of 1 input pixel of noise. We used standard neural network training techniques to find the weights and biases which we mapped to the critical current in the MJJ elements. The total spiking energy for the computation of one input is roughly 2 aJ (≈ 2fJ with refrigeration), which is comparable to the energy of a single human synaptic event ≈ 10 fJ. To realize this potentially ultralow power computational scheme, we need to significantly reduce or eliminate the need for the stabilization resistor, which currently dominates the power consumption of the circuit. Given, the scalability of neural nets, we expect that large images can be processed in ns time scales using this technology.

Acknowledgment

The authors thank Manuel Castellanos Beltran for helpful circuit design discussions and the IARPA C3 program for partial support.

References

[1] Josephson, B.D.: "Possible new effects in superconductive tunnelling", Physics Letters, 1962, 1, (7), pp. 251-253

[2] Tinkham, M.: 'Introduction to superconductivity' (Dover, 1996. 1996)

[3] Chen, W., Rylyakov, A.V., Patel, V., Lukens, J.E., and Likharev, K.K.: 'Rapid Single Flux Quantum T-flip flop operating up to 770 GHz', Ieee Transactions on Applied Superconductivity, 1999, 9, (2), pp. 3212-3215

[4] Schneider, M.L., Donnelly, C.A., Russek, S.E., Baek, B., Pufall, M.R., Hopkins, P.F., Dresselhaus, P.D., Benz, S.P., and Rippard, W.H.: 'Ultra-low power artificial synapses using nano-textured magnetic Josephson junctions', Submitted, 2017

[5] Holmes, S., Ripple, A.L., and Manheimer, M.A.: 'Energy-Efficient Superconducting Computing-Power Budgets and Requirements', Ieee Transactions on Applied Superconductivity, 2013, 23, (3), pp. 10

[6] Segall, K., Guo, S.Y., Crotty, P., Schult, D., and Miller, M.: 'Phase-flip bifurcation in a coupled Josephson junction neuron system', Physica B, 2014, 455, pp. 71-75

[7] Hirose, T., Asai, T., and Amemiya, Y.: 'Pulsed neural networks consisting of single-flux-quantum spiking neurons', Physica C, 2007, 463, pp. 1072-1075

[8] Onomi, T., Kondo, T., and Nakajima, K.: 'High-speed single flux-quantum up/down counter for neural computation using stochastic logic', in Hoste, S., and Ausloos, M. (Eds.): '8th European Conference on Applied Superconductivity' (Iop Publishing Ltd, 2008)

[9] Yamanashi, Y., Umeda, K., and Yoshikawa, N.: 'Pseudo Sigmoid Function Generator for a Superconductive Neural Network', IEEE Transactions on Applied Superconductivity, 2013, 23, (3), pp. 1701004-1701004

[10] 'Whiteley Research Inc., WRSPICE, http://www.wrcad.com/wrspice.html'

[11] Van Duzer, T., and Turner, C.W.: 'Superconductive Devices and Circuits' (Prentice Hall, 1999. 1999)

[12] Russek, S.E., Donnelly, C.A., Schneider, M.L., Baek, B., Pufall, M.R., Rippard, W.H., Hopkins, P.F., Dresselhaus, P.D., and Benz, S.P.: 'Stochastic single flux quantum neuromorphic computing using magnetically tunable Josephson junctions', Proc. IEEE ICRC, 2016, pp. 1-5

[13] Kirichenko, D.E., Sarwana, S., and Kirichenko, A.F.: 'Zero Static Power Dissipation Biasing of RSFQ Circuits', IEEE Transactions on Applied Superconductivity, 2011, 21, (3), pp. 776-779

[14] Prezioso, M., Merrikh-Bayat, F., Hoskins, B.D., Adam, G.C., Likharev, K.K., and Strukov, D.B.: 'Training andoperation of an integrated neuromorphic network based on metal-oxide memristors', Nature, 2015, 521, (7550), pp. 61-64

[15] Nielsen, M.A.: 'Neural Networks and Deep Learning' (Determination Press, 2015. 2

Baek, Burm; Donnelly, Christine; Hopkins, Peter; Pufall, Matthew; Rippard, William; Russek, Stephen; Schneider, Michael.
"Energy efficient Single Flux Quantum Based Neuromorphic Computing."
Paper presented at 2017 IEEE International Conference on Rebooting Computing (ICRC) , Washington, DC, United States. November 8, 2017 - November 9, 2017.

Aperture Arrays for Subnanometer Calibration of Optical Microscopes

C. R. Copeland[1, 2], C. D. McGray[3, 4], J. Geist[3], J. A. Liddle[1], B. R. Ilic[1], and S. M. Stavis[1,*]

[1]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899

[2]Maryland Nanocenter, University of Maryland, College Park, Maryland 20742,

[3]Engineering Physics Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899

[4]Modern Microsystems, Silver Spring, Maryland 20899

[*]samuel.stavis@nist.gov

***Abstract*—We fabricate and test subresolution aperture arrays as calibration devices for optical localization microscopy. An array pitch with a relative uncertainty of approximately three parts in ten thousand enables localization with subnanometer accuracy.**

Keywords—aperture; subnanometer; localization; microscopy

I. Introduction

Optical microscopy methods of imaging and localizing subresolution emitters are having impact in diverse studies ranging from microelectromechanical to biological systems [1–4]. Such measurements can have subnanometer precision, even for single-molecule emitters [5], but if calibrations ensuring accuracy at corresponding length scales are absent, then this can be false precision. There has been commercial interest in developing subresolution aperture arrays as calibration devices for optical microscopes [6], as well as recent application of such devices to localization microscopy in three dimensions [7]. However, aperture arrays enabling measurements that are traceable to the SI at subnanometer scales are not yet in common use. Here, we fabricate and test such nanostructures, introducing a practical approach to optical localization that is accurate at subnanometer scales.

II. Methods and Materials

A. Device Fabrication

We fabricate aperture arrays in a platinum film on a silica substrate. A film thickness of approximately 80 nm results in low background noise from light transmission, and a substrate thickness of approximately 0.17 mm enables the future calibration of objective lenses with oil immersion for detection of single fluorophores. We use electron-beam lithography and ion milling to pattern sub-resolution apertures on a pitch of 10 μm, defining the critical dimension of our calibration device. Our lithography system positions the electron beam with a nominal accuracy of 2 nm, which we take as an initial estimate of the standard deviation of aperture placement in one lateral dimension. The uncertainty of the array pitch is greater than this value by a factor of $\sqrt{2}$, giving a relative uncertainty of the array pitch of approximately 3×10^{-4}. We will revisit this analysis in a future study. The extent of the aperture array exceeds the field of our optical microscope.

The authors acknowledge support of this research under the National Institute of Standards and Technology (NIST) Innovations in Measurement Science Program, the NIST Center for Nanoscale Science and Technology, and the NIST Physical Measurement Laboratory. C. R. C. acknowledges support under the Cooperative Research Agreement between the University of Maryland and the NIST Center for Nanoscale Science and Technology, award number 70ANB10H193, through the University of Maryland.

B. Optical Microscopy

A light emitting diode trans-illuminates the aperture array at a wavelength of approximately 630 nm. An objective lens with corrections for chromatic and flatfield aberrations, a nominal magnification of 50×, and a numerical aperture of 0.55 collects light transmitted through the apertures. A tube lens projects the image of the aperture array onto a complementary metal-oxide-semiconductor (CMOS) camera with 2048×2048 pixels with a nominal size of 6.5 μm. The camera records optical micrographs at a video rate of 10 Hz. Fig. 1 shows a representative optical micrograph of a small region of the aperture array. Each subresolution aperture appears as the point spread function of the imaging system. We use weighted least-squares estimation to fit each point spread function to a symmetric Gaussian model. Our localization analysis accounts for the gain and noise of each pixel of the CMOS camera over its full field. We will describe the details of our localization analysis in a future study. In each micrograph, the localization precision for each aperture is less than 1 nm.

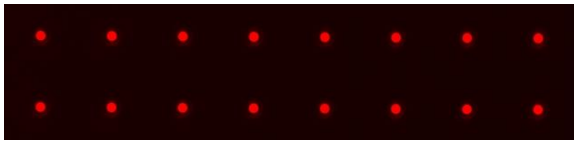


Fig. 1. Optical micrograph showing an array of superresolution apertures in a platinum film on a silica substrate. The array pitch is 10 μm with a relative uncertainty of 3×10^{-4}. Aperture localization enables microscope calibration.

III. Results and Discussion

A. Magnification Calibration

Calculation of distances between apertures across a micrograph enables mapping of the image pixel size. Fig. 2 shows such a map, with linear interpolation between apertures. A radial pattern is evident, possibly from lens manufacture. The image pixel size varies from 127.10 nm to 127.43 nm. In contrast, the nominal value of the image pixel size is 130 nm, resulting in a range of errors of 2 % or more. Many previous studies have used relatively rudimentary methods to calibrate image pixel size, potentially undermining the accuracy of otherwise precise localization measurements.

Copeland, Craig; Geist, Jon; Ilic, Bojan; Liddle, James; McGray, Craig; Stavis, Samuel.
"Aperture Arrays for Subnanometer Calibration of Optical Microscopes."
Paper presented at International Conference on Optical MEMS and Nanophotonics, Santa Fe, NM, United States. August 13, 2017 - August 17, 2017.

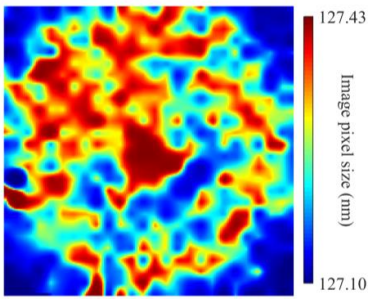


Fig. 2. Spatial map showing variation of image pixel size, possibly from lens manufacture, over the full field of the imaging system. The nominal value of image pixel size is 130 nm, corresponding to errors of 2 % or more.

B. Noise Analysis

We assume that the aperture array is mechanically stable, allowing analytical elimination of any lateral motion that is common between apertures. This isolates the apparent lateral motion of each aperture due to photon shot noise, which is the physical limit of uncertainty in the localization of subresolution emitters [8, 9]. Temporal averaging of photon shot noise reduces the Allan deviation for single apertures through the subnanometer scale and into the picometer scale. Fig. 3 shows a representative analysis for a single lateral dimension. The black line is the mean value and the gray bound denotes the standard deviation of the Allan deviation of 50 apertures. The slope of approximately -0.5 is consistent with the inverse square root of photon count, which increases linearly with averaging interval. Uncertainty in the size of image pixels from the relative uncertainty of the array pitch sets an inaccuracy floor of approximately 3×10^{-2} nm, which we approach in an averaging interval of less than 1 min. Much of this precision would be false precision in the absence of the preceding magnification calibration.

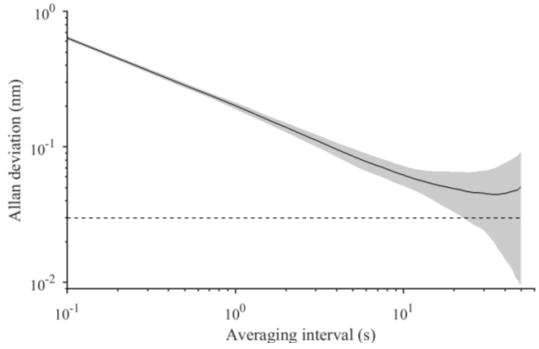


Fig. 3. Temporal averaging of photon shot noise reduces the mean value of Allan deviation of aperture location toward an inaccuracy floor of 3×10^{-2} nm. The gray bound denotes the standard deviation of 50 apertures.

IV. Conclusions

We are concerned that the increasing interest in achieving localization precision at subnanometer scales has advanced ahead of the corresponding metrology foundation for ensuring localization accuracy. Here, we have introduced a practical approach to achieve both accuracy and precision in optical localization microscopy extending below the subnanometer scale. In a future study, we will revisit the various topics that we have noted here, and we will apply our new measurement capability to quantify any motion of fluorescent nanoparticles adsorbed to imaging substrates. While many previous studies have assumed that this arrangement results in nominally static fiducials, there are open questions in the literature on this topic [9, 10] that we can now answer more definitively.

References

[1] C. D. McGray, S. M. Stavis, J. Giltinan, E. Eastman, S. Firebaugh, J. Piepmeier, J. Geist, M. Gaitan, "MEMS kinematics by super-resolution fluorescence microscopy," Journal of Microelectromechanical Systems, 22, 2, 115-123, 2013.

[2] C. R. Copeland, C. D. McGray, J. Geist, V. A. Aksyuk, S. M. Stavis, "Characterization of electrothermal actuation with nanometer and microradian precision," Transducers 2015, 792-795, 2015.

[3] C. R. Copeland, C. D. McGray, J. Geist, V. A. Aksyuk, S. M. Stavis, "Transfer of motion through a microelectromechanical linkage at nanometer and microradian scales," Microsystems & Nanoengineering, 2, 16055, 2016.

[4] P. P. Mathai, J. A. Liddle, S. M. Stavis, "Optical tracking of nanoscale particles in microscale environments," Applied Physics Reviews, 3, 011105, 2016.

[5] A. Pertsinidis, Y. Zhang, S. Chu. "Subnanometre single-molecule localization, registration and distance measurements," Nature, 466, 647-51, 2010.

[6] T. Matsuzawa, G. Ryu, Y. Eda, T. Morita. "Lens evaluation device," United States Patent, US7747101 B2, 2010.

[7] A. V. Diezmann, M. Y. Lee, M. D. Lew, W. E. Moerner, "Correcting field-dependent aberrations with nanoscale accuracy in three-dimensional single-molecule localization microscopy," Optica, 2, 985-993, 2015.

[8] C. D. McGray, C. R. Copeland, S. M. Stavis, J. Geist, "Centroid precision and orientation precision of planar localization microscopy," Journal of Microscopy, 263, 3, 238-249, 2016.

[9] K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, "Optimized localization analysis for single molecule tracking and super resolution microscopy," Nature Methods, 7, 377-381, 2010.

[10] A. R. Carter, G. M. King, T. A. Ulrich, W. Halsey, D. Alchenberger, T. T. Perkins, "Stabilization of an optical microscope to 0.1 nm in three dimensions," Applied Optics, 46, 3, 421-427, 2007.

Copeland, Craig; Geist, Jon; Ilic, Bojan; Liddle, James; McGray, Craig; Stavis, Samuel.
"Aperture Arrays for Subnanometer Calibration of Optical Microscopes."
Paper presented at International Conference on Optical MEMS and Nanophotonics, Santa Fe, NM, United States. August 13, 2017 - August 17, 2017.

SI-Traceable Top-of-the-Atmosphere Lunar Irradiance: Potential Tie-Points to the Output of the ROLO Model

Steven W. Brown *[a], Robert E. Eplee, Jr [b], Xiaoxiong Xiong[c]
[a]National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899;
[b]Science Applications International Corporation, 4600 Powdermill Road, Suite 400, Beltsville, MD 20705;
[c]Science and Exploration Directorate, National Aeronautics and Space Administration's Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, MD 20771.

ABSTRACT

The United States Geological Survey (USGS) has developed an empirical model, known as the Robotic Lunar Observatory (ROLO) Model, that predicts the reflectance of the Moon for any Sun-sensor-Moon configuration over the spectral range from 350 nm to 2500 nm. The lunar irradiance can be predicted from the modeled lunar reflectance using a spectrum of the incident solar irradiance. While extremely successful as a relative exo-atmospheric calibration target, the ROLO Model is not SI-traceable and has estimated uncertainties too large for the Moon to be used as an absolute celestial calibration target.

In this work, two recent absolute, low uncertainty, SI-traceable top-of-the-atmosphere (TOA) lunar irradiances, measured over the spectral range from 380 nm to 1040 nm, at lunar phase angles of 6.6° and 16.9°, are used as tie-points to the output of the ROLO Model. Combined with empirically derived phase and libration corrections to the output of the ROLO Model and uncertainty estimates in those corrections, the measurements enable development of a corrected TOA lunar irradiance model and its uncertainty budget for phase angles between ±80° and libration angles from 7° to 51°. The uncertainties in the empirically corrected output from the ROLO model are approximately 1 % from 440 nm to 865 nm and increase to almost 3 % at 412 nm. The dominant components in the uncertainty budget are the uncertainty in the absolute TOA lunar irradiance and the uncertainty in the fit to the phase correction from the output of the ROLO model.

Keywords: calibration, lunar model, moon

1. INTRODUCTION

Identifying, quantifying and differentiating between natural and anthropogenic changes to the Earth system are fundamental objectives of many Earth remote sensing programs. Quantification of changes in climate-relevant global variables has been a driving force behind increasingly stringent uncertainty requirements on measurements made by satellite sensors. The international, multi-satellite, multi-sensor nature of the task underscores the need for traceability of those measurements to the International System of Units (SI). Radiometric sensors degrade on-orbit in non-deterministic ways due to the harsh environment of Space. Consequently, on-orbit sensor characterization and calibration methodologies have been developed to track responsivity changes of satellite sensors.

A primary component of many Earth observing sensors' on-orbit calibration strategy is making consistent measurements of the Moon. The Moon exhibits radiometric characteristics that make it an excellent exo-atmospheric radiometric target for Earth-viewing remote sensing sensors: the surface of the Moon is exceptionally stable, that is, its reflectance does not degrade in time; the lunar radiance or irradiance in the reflective solar region has a magnitude similar to Earth scenes; and views of the Moon are possible for satellite sensors viewing the Earth in any Earth orbit [1, 2]. Because the lunar reflectance is temporally stable, using the Moon as a radiometric target provides an avenue for instrument cross-comparison without the need to look through the Earth's atmosphere or to view a common target contemporaneously – as long as the input solar irradiance is stable. Finally, because the reflectance of the Moon is

Brown, Steven; Eplee, Robert; Xiong, Xiaoxiong.
"Si-traceable top-of-the-atmosphere lunar irradiance: Potential tie-points to the output of the ROLO Model."
Paper presented at SPIE, San Diego, CA, United States. August 6, 2017 - August 11, 2017.

temporally stable, measurements of the Moon by on-orbit sensors can be reprocessed as modeling and knowledge of the absolute lunar irradiance improves, increasing the historical knowledge of an on-orbit sensor's radiometric characteristics. Note that while the reflectance of the Moon is invariant over time scales of interest, the lunar irradiance itself is a function of phase and libration angles as well as the Earth-Moon-Sun distance.

The potential of the Moon as an exo-atmospheric satellite sensor calibration source motivated the Unites States Geological Survey (USGS) and Northern Arizona University, with funding by the National Aeronautics and Space Administration (NASA), to develop the ROLO Observatory at the USGS facility in Flagstaff, AZ [3], measure the lunar irradiance as a function of phase and libration angles, and develop a Model to facilitate its use as an absolute celestial radiometric calibration source [2, 4, 5]. The ROLO instrument is a dual-telescope, multi-band filter radiometer with a total of 32 user-selectable filters spanning the spectral region from 350 nm to 2450 nm with bandpasses ranging from ≈10 nm to ≈60 nm [2]. The ROLO Model is a reflectance model developed using measurements from ROLO's 32 filter bands. ROLO provided the USGS with a continuous 6+-year time series of radiometric observations of the Moon over a subset of phase and libration angles from which the disk-integrated lunar irradiance or the spatially resolved lunar radiance could be derived [6]. Using the disk-integrated irradiance in conjunction with the Wehrli model of solar irradiance [7], the USGS developed the ROLO Model, an analytical lunar reflectance model, with continuous variables of lunar phase and libration [2, 8].

The ROLO Model has had several versions over the years as it continues to be refined. The current version of the ROLO Model, 3.11g, is used in this work [2]; it has been the disseminated version of the Model since 2005 and is the version currently used by NASA's Ocean Biology Processing Group (OBPG) [9], responsible for SeaWiFS, MODIS and VIIRS instrument ocean color data processing, and by international organizations such as the Global Space-based Inter-Calibration System (GSICS) and the Committee on Earth Observation Satellites (CEOS) in their implementation of the ROLO Model, the GSICS implementation of the ROLO model (GIRO model) [10].

The ROLO Model combined with satellite sensor measurements of the Moon have been used successfully to trend the temporal degradation of satellite sensor channel responsivities; the canonical success story is SeaWiFS, which achieved a relative uncertainty in the TOA lunar irradiance of 0.13 % by comparing monthly lunar views with the ROLO Model over the 13-year life of the mission [11]. To achieve this level of precision, the lunar measurements were taken at nominal phase angles of $+7^o$ (from $+5^o$ to $+10^o$) and -7^o (from -6^o to -8^o), resulting in 90 observations at a nominal phase angle of -7^o and 55 observations at $+7^o$ phase. The data were corrected for lunar libration and detector focal plane temperature [11, 12]. The success of the ROLO Model for sensor response trending is reflected in the number of Earth-observing instruments looking at the Moon and utilizing the ROLO Model. U. S. instruments such as SeaWiFS; MODIS on NASA's Aqua and Terra satellites; the Hyperion hyperspectral imaging spectrometer on NASA's Earth Observing-1 satellite; the Multi-Angle Imaging SpectroRadiometer (MISR) on Terra; the Operational Land Imager (OLI) on Landsat 8; the Advanced Baseline Imager (ABI) on the joint National Oceanic and Atmospheric Administration (NOAA)/NASA Geostationary Operational Environmental Satellites – R Series (GOES-R), now GOES-16; and the Suomi National Polar-orbiting Partnership (SNPP) Visible Infrared Imaging Radiometer Suite (VIIRS) have all used the Moon as a celestial radiometric target. Future missions will also view the Moon: lunar measurements are planned for the Joint Polar Satellite System (JPSS) J1 VIIRS; the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) Mission; and the Climate Absolute Radiance and Refractivity Observatory (CLARREO) Mission.

In developing the ROLO Model, extinction measurements of stars were used to develop a time-dependent characterization of the atmosphere, which is then used to determine the extinction in the ROLO lunar images [2]. The ROLO telescopes were calibrated against the standard star Vega (α-Lyr) [6] using flux measurements by Hayes [13] at 555.6 nm scaled to a model stellar spectrum developed by Castelli and Kurucz [14]. The scaled, modeled Vega spectrum was convolved with the ROLO instrument bandpasses, giving band-averaged values for each of the 32 filter bands. These inputs along with reflectance measurements of returned Apollo lunar samples served as the foundation of the ROLO Model [2, 8]. The measurement uncertainty in the Vega flux at 555.6 nm was estimated by Hayes to be 1.5 %. The random error in scaling to the model stellar spectrum developed by Castelli and Kurucz [14] is estimated to be ≈1 % or less. Taking the root-sum-square of the two uncertainty components, a rough estimate of the minimum uncertainty in the absolute flux from Vega in the visible-near infrared (visNIR) spectral range is 1.8 %. More recently, model assumptions that Vega is a slow to moderately rotating star has been called into question [15]. The current hypothesis is that Vega is a rapidly rotating star seen pole-on [15]. The new model leads to a prediction of unusual

Brown, Steven; Eplee, Robert; Xiong, Xiaoxiong.
"Si-traceable top-of-the-atmosphere lunar irradiance: Potential tie-points to the output of the ROLO Model."
Paper presented at SPIE, San Diego, CA, United States. August 6, 2017 - August 11, 2017.

brightness and excess near-infrared emission compared to visible emission, leading to a higher uncertainty in the modelled irradiance from Vega. As discussed by Kieffer and Stone [2, 8], there are additional uncertainties in the ROLO Model from the lunar reflectance, image processing, atmospheric correction, phase dependence, etc. that add several percent to the uncertainty in the model, leading to the current uncertainty estimate for the ROLO Model in the visNIR spectral range being between 5 % and 10 % [8].

No single organization or country can independently develop the Earth remote sensing suite of instruments required for global, community relevant measurements. Organizations such as CEOS [16] and GSICS [17] have been developed to foster collaboration between stakeholders (nations and national remote sensing institutions, *e.g.*, NASA, NOAA and the USGS in the U. S.) in global Earth remote sensing instruments, measurements, and data products. Measurement traceability to the SI is crucial for the successful merging of global Earth remote sensing data sets in that traceability enables measurements by different sensors to be compared with confidence. Typical at-sensor uncertainty requirements for radiometric measurements range from 0.5 % for ocean color, to 2 % for vegetation and 3 % for aerosols; stability requirements are stricter: 0.1 % for ocean color, 0.8 % for vegetation, and 1.5 % for aerosols [18-20]. The uncertainty requirements for future missions like CLARREO are even more stringent: 0.3 % ($k = 2$) in the reflected solar region [21, 22]. While extremely successful as a relative on-orbit target, the Moon is not currently useful as an absolute celestial target because the ROLO Model is not traceable to the SI and the estimated uncertainties in the Model are too large compared with current requirements.

Recently, two measurements of the absolute, SI-traceable measurements of the top-of-the-atmosphere (TOA) lunar irradiance at phase angles of 6.6^{o} and 16.9^{o} were made by researchers from the National Institute of Standards and Technology (NIST) using time series of visible-near infrared (visNIR) lunar irradiance measurements at the Fred Lawrence Whipple Observatory, Mt. Hopkins, Arizona [23]. The ground-based measurements were propagated to the TOA using Langley analysis to account for atmospheric attenuation. The measurements are used to tie the output of the ROLO Model to the absolute, SI-traceable, TOA lunar irradiance at a phase angle of 6.6^{o}. The NIST TOA lunar irradiances were band-averaged using the National Aeronautics and Space Administration's (NASA's) Sea-viewing Wide Field-of-view Sensor (SeaWiFS) bands, Table 1 [24], and compared with the predicted lunar irradiance using the ROLO Model, v. 3.11g, combined with a model of the solar spectral irradiance.

Using the absolute TOA lunar irradiance measurements to scale the output of the ROLO model, additional corrections as functions of lunar phase and libration angles are necessary to correct the output from the ROLO Model for different Sun-sensor-Moon configurations. Data from NASA's SeaWiFS instrument and data from France's National Centre for Space Studies's (CNES's) PLEIADES instruments, PLEIADES-1A and 1B, are used to derive an empirical phase correction to the output of the ROLO Model. SeaWiFs data are used to develop an empirical correction to the libration dependence at low phase angles; a data set from the SUOMI-National Polar-orbiting Partnership (S-NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) is used to develop an empirical correction to the libration dependence at high phase angles. Band-center wavelengths and bandwidths of PLEIADES [25] and VIIRS instruments [26] are listed in Table 1. Correction factors and associated uncertainties to the ROLO Model-predicted lunar irradiance are presented. Implications of the empirical correction to the output from the ROLO Model for transitioning the Moon toward an absolute, exo-atmospheric, SI-traceable satellite sensor calibration source with uncertainties commensurate with science data product requirements are discussed.

2. ABSOLUTE SI-TRACEABLE EXO-ATMOSPHERIC LUNAR IRRADIANCE

In 2008, NIST began a series of observation field campaigns at the Whipple Observatory on Mt. Hopkins in southern Arizona with the goal of establishing stellar and lunar radiometric source standards with uncertainties less than 1 % to support requirements of the astronomical and Earth remote sensing communities. Measurements of the irradiance from the standard reference star Vega at the original calibration wavelength of 555.6 nm was one of NIST's primary stellar calibration goals and it led to the use of a fiber-coupled spectrograph with a silicon-based CCD detector being mounted on the telescope. The spectrograph has a spectral range from 380 nm to 1040 nm and a resolution of approximately 4 nm [27]. In 2011, observations at Mt. Hopkins were extended to include the Moon [23]. Lunar irradiance measurements were made with a 106 mm refracting telescope with a 50.8 mm diameter integrating sphere

Brown, Steven; Eplee, Robert; Xiong, Xiaoxiong.
"Si-traceable top-of-the-atmosphere lunar irradiance: Potential tie-points to the output of the ROLO Model."
Paper presented at SPIE, San Diego, CA, United States. August 6, 2017 - August 11, 2017.

Table 1. Band-center wavelengths and FWHM bandwidths of (a) SeaWiFS, (b) PLEIADES and (c) VIIRS bands.

(a)	SeaWiFS		(b)	PLEIADES		(c)	VIIRS	
Band	Band Center Wavelength [nm]	Bandwidth [nm]	Band	Band Center Wavelength [nm]	Bandwidth [nm]	Band	Band Center Wavelength [nm]	Bandwidth [nm]
1	412.6	20.3	B0	490	120	1	412	20
2	443.5	20	B1	550	120	2	445	18
3	490.8	20.7	B2	660	120	3	488	20
4	509.9	22.3	B3	850	200	4	555	20
5	554.6	18.3				5	672	20
6	668.3	19.8				6	746	15
7	765	41.4				7	865	39
8	866.2	41.3						

placed at its focal plane where the imaged lunar disc is approximately one-third the size of the telescope integrating sphere's 12 mm entrance aperture. The integrating sphere spatially scrambled light from the entire lunar disc, uniformly illuminating a flexible light guide that coupled incident flux from the integrating sphere into a spectrograph; the same model spectrograph was used for both stellar and lunar measurements.

The telescope was calibrated by observing an "artificial moon" light source with a known spectral irradiance, shown schematically in Figure 1. The artificial moon is a 300 mm-diameter integrating sphere (SIS) placed 36.42(3) m from the telescope's effective aperture (D_2) and illuminated with a quartz-tungsten-halogen (QTH) lamp controlled by the Lamp Power Supply (LPS). The source irradiance is determined using a reference spectrograph (R-SG) with its defining aperture located a distance D_1 from the exit aperture of the SIS. Atmospheric extinction is negligible over this horizontal path excepting molecular absorption bands that are not considered in this work. The distance between the calibration

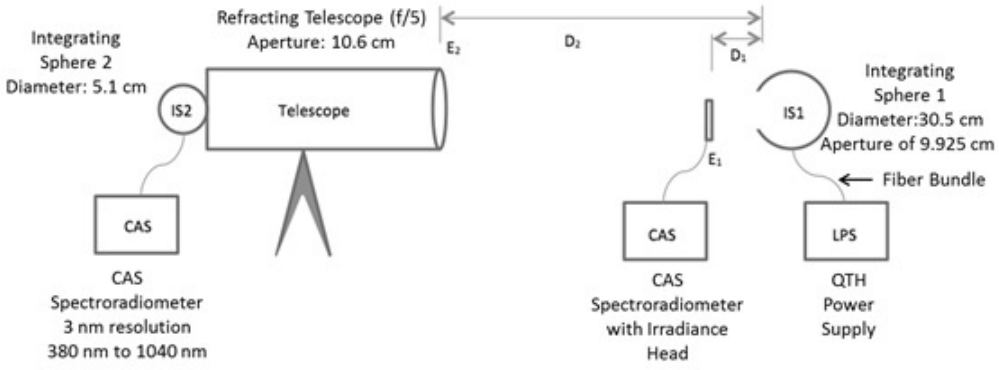


Figure 1. Experimental Setup used to calibrate the telescope: SIS = Spherical Integrating Sphere; LPS = Lamp Power Supply; R-SG = Reference Spectrograph, with Irradiance Head E1 located a distance D1 from the SIS exit aperture; D2 is the separation between the exit aperture of the SIS and the entrance aperture of the telescope; DIS = the Diffuse Integrating Sphere; and T-SG= the Telescope Spectrograph.

source and the telescope entrance aperture and the lamp current are chosen to give the imaged sphere aperture a smaller size and similar brightness to the full Moon when viewed with the telescope. The irradiance at the telescope reference plane is given by Eq. 1 and the telescope responsivity for array element *i* is given by the Signal divided by the Irradiance, Eq. 2, where *i*=1 to 1024, the number of elements in the spectrograph detector array:

$$E_2 = \left(D_1/D_2\right)^2 E_1 \tag{1}$$

and

$$R_i = S_i/E_{2i} \ . \tag{2}$$

Repeating the calibration every two hours throughout the night established that the telescope calibration was stable over the course of a night to better than 0.2 % from 420 nm to 1000 nm.

The absolute TOA lunar spectral irradiance, derived from a time series of ground-based lunar measurements corrected for transmittance through the atmosphere by the Beer-Lambert-Bougher Law on the evenings of November 28 and 29, 2012, scaled to measurement dates/UT times of 29Nov2012/03:29:07UT and 30Nov2012, 02:11:07UT, are presented in Figure 2 [23]. Lunar spectra were recorded in 1 minute intervals over the course of several hours to allow for an accurate measurement of atmospheric transmittance. The time stamp on each lunar irradiance spectrum was tied to a GPS receiver and is accurate to within 200 μs. The Sun-Moon-Observer angle, that is, the lunar phase angle, changes over the course of a night's measurements. On 29 November 2012, for example, the lunar phase changed from 17^o to 20^o over the course of the measurements. The changing phase angle produces a TOA lunar irradiance that changes throughout the night as well. The ROLO Model itself was used to correct for the phase change in TOA lunar irradiance through the night. The lunar irradiances were scaled to phase angles of 6.6^o and 16.9^o, respectively. Figure 2 shows the absolute TOA lunar spectral irradiance for the two nights after correcting for ozone, aerosols, Rayleigh scattering and molecular absorption. Solar-generated Fraunhofer lines are identified in the figure. The lunar phase changed approximately 10^o between the two sets of measurements while the magnitude of the lunar irradiance changed by approximately 40 %. They were the only 2 nights in 40 nights of observations over 2 years where the atmosphere was stable enough to produce the high-quality Langley plots needed to correct for atmospheric attenuation.

The measurement chain back to primary national standards was established by measuring the spectral irradiance of the artificial moon with a NIST-calibrated SI-traceable reference spectrograph during every telescope calibration. The reference spectrograph's spectral irradiance responsivity is tied to the SI through measurements of a reference standard FEL-type lamp of known spectral irradiance [28]. The spectrograph was characterized for stray light and a correction matrix was developed [29] and applied to each measurement. The spectrograph measured the two FEL-type standard irradiance lamps at NIST before and after each deployment to the Mt. Hopkins site. To establish traceability to the SI, an uncertainty budget was developed. Figure 3 is a graphical presentation of the combined standard uncertainty budget for TOA lunar irradiance measurements on November 29, 2012 (UTC) and November 30, 2012 (insert) [23]. The uncertainty budgets for the two nights are similar, with the uncertainty slightly higher for the November 30 data set in the 550 nm to 650 nm spectral region due to increased ozone and possibly aerosol absorption. Three dominant uncertainty components in the budget are: the uncertainty in the fit to the Beer-Lambert-Bouger Law, the uncertainty in the correction for atmospheric absorption, and the uncertainty in the calibration of the telescope. The dominant measurement uncertainty component is the telescope calibration; the uncertainty budget for the telescope calibration is, in turn, dominated by the uncertainty in the responsivity of the reference spectrograph [23].

3. DEVELOPMENT OF AN EMPIRICAL CORRECTION TO THE OUTPUT FROM THE ROLO MODEL

In Section 3.1, the absolute TOA lunar irradiance measured at Mt. Hopkins is compared with the output from the ROLO Model using the SeaWiFS ocean color bands for a spectral comparison. The uncertainty budgets for the two TOA lunar irradiance spectra are presented. The output of the ROLO Model can be tied to the absolute measurements;

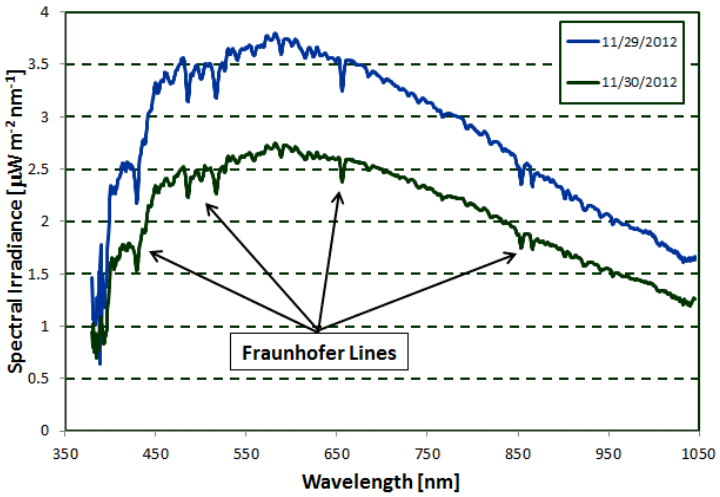


Figure 2. Absolute TOA Lunar spectral irradiance measured at the Whipple Observatory, Mt. Hopkins AZ on November 29, 2012 and November 30, 2012 with atmospheric absorption features removed.

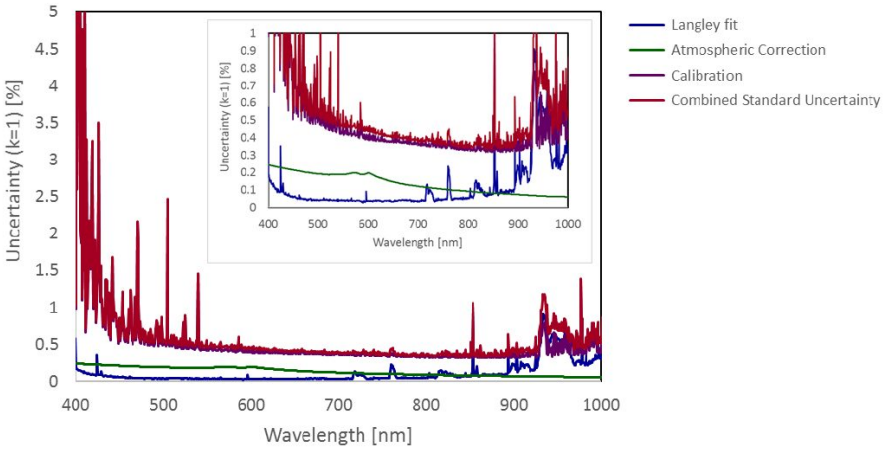


Figure 3. Lunar irradiance uncertainty budget for Nov. 29th and Nov. 30th (inset). Blue: Langley Fit; Green: Ozone and Aerosols; Purple: Telescope calibration; Red: Combined Standard (k=1) Uncertainty.

residual phase and libration corrections, and their uncertainties, are required for a generalized correction to the output of the Model. The phase and libration corrections to the output of the ROLO model presented in this work are based on empirical data from SeaWiFS, PLEIADES and VIIRS satellite sensors. Corrections are applied to the output of the ROLO Model for any phase and libration angle and the uncertainties in their corrections are estimated in Sections 3.2 and 3.3, respectively. In Section 3.4, an uncertainty budget is presented for the empirically corrected output of the ROLO Model.

3.1 Comparison between the absolute TOA lunar irradiance measured at Mt. Hopkins and output from the ROLO Model using SeaWiFS ocean color sensor bands

The modeled and measured TOA lunar irradiances have been convolved with SeaWiFS bands using Eq. 3 for band-averaged quantities:

$$E_{x,i}^{BA} = \frac{\int \left(E_x(\lambda) \cdot R_i(\lambda) \right) d\lambda}{\int R_i(\lambda) d\lambda} . \quad (3)$$

$E_{x,i}^{BA}$ is the band-averaged lunar irradiance based on NIST measurements or predicted by the ROLO Model and $R_i(\lambda)$ is the relative spectral response of the SeaWiFS bands, *i*=1 to 8. Band-center wavelengths and bandpasses for SeaWiFS are given in Table 1; the integral is over the total band region, from approximately 380 nm to 1000 nm. The comparisons are between the TOA Lunar irradiances derived from NIST measurements and the USGS model-predicted Lunar irradiances and do not involve the absolute calibration of SeaWiFS.

The SeaWiFS band-averaged absolute TOA spectral irradiances from November 29th and 30th, 2012 (UT) are shown in Figure 4. In agreement with the spectral measurements shown in Figure 2, the band-averaged lunar irradiance changed by approximately 40 % between the 2 nights. The resulting comparisons of the ratio between the NIST-measured to the ROLO Model-predicted SeaWiFS band–averaged irradiances for the two nights are shown in Figure 5. These comparisons show the relative biases between the NIST-measured absolute lunar irradiances and the ROLO Model's predicted absolute lunar irradiance scale. The level of agreement in the ratio between the two nights of observations ranges from 1 % for bands 1 and 2; to 0.3 % for bands 3, 6 and 8; to better than 0.1 % for bands 4, 5, and 7; well within the NIST measurement uncertainties shown in Figure 3.

3.2 Phase correction to the output of the ROLO Model

There is a strong sensor dependence to biases in comparisons with the ROLO Model [30], and wavelength dependences for a particular sensor [31]. In developing a correction to the phase dependence of the output of the ROLO Model, only data from sensors that made measurements at a phase angle of +7°, where absolute measurements were made, were used; instruments considered in this work were SeaWiFS and PLEIADES-1A and -1B sensors. Differences between the instrument and the ROLO Model are set to 0 at a phase angle +7°.

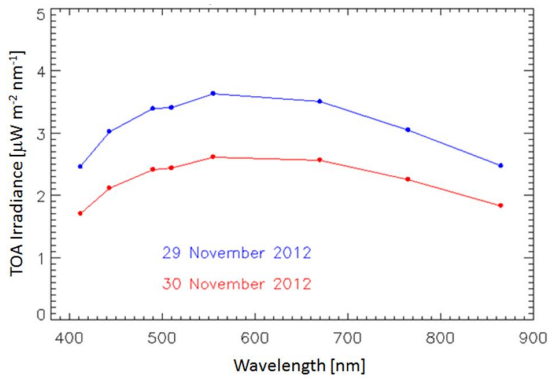


Figure 4. The TOA lunar irradiance band-averaged over the SeaWiFS bands.

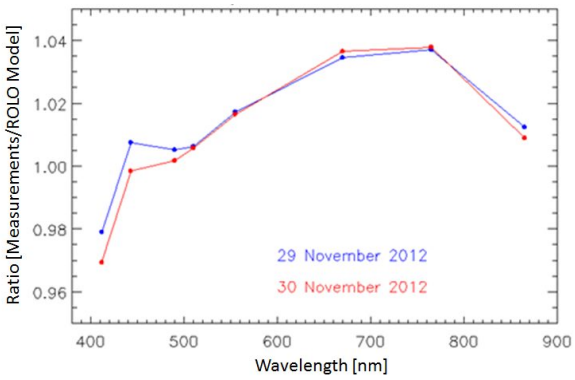


Figure 5. The NIST–measured lunar irradiance ratioed to ROLO Model-predicted irradiance. Both the measured and the model-predicted irradiances are band-averaged to the SeaWiFS bands. The error bars are the SeaWiFS band-averaged, combined standard (k = 1) uncertainty in the NIST-measured TOA irradiance.

SeaWiFS made 90 measurements at a nominal phase angle of -7° ± 1°, 44 measurements at a nominal phase angle of +7° ± 1°, 27 measurements at angles ranging from -27° to -49°, and 32 measurements from +27° to +66° [11, 32]. Under normal operations, PLEIADES instruments image the Moon twice a month at a phase angle close to 40°. However, an intensive series of lunar observations, called the Pleiades Orbital Lunar Observations (POLO), acquired approximately 1000 observations of the Moon covering the phase angles from +115° to -115° [31]. Lunar images were acquired as often as every orbit (≈100 minutes) during the POLO campaign.

Figure 6a shows the ratio between the lunar measurement and the ROLO Model as a function of phase angle and wavelength for (left) SeaWiFS and (right) PLEIADES sensors. A phase angle-dependent difference in the ratio between the lunar measurements and the ROLO Model was observed, as was a spectrally dependent bias. A phase correction to the ROLO Model is developed using the SeaWiFS and PLEIADES data sets by setting the difference between the instrument measurement and the ROLO Model to 0 at +7° phase, the phase angle of one of the two absolute measurements from Mt. Hopkins, Figure 6b. Error bars for the SeaWiFS data are shown in the figure; no uncertainties were provided with the PLEIADES data.

For SeaWiFS, all spectral measurements fall within the combined k = 1 uncertainties and there is no discernable spectral dependence to the phase correction. The PLEIADES measurements show less spectral variation than the SeaWiFS measurements. Assuming the PLEIADES uncertainties are similar to the SeaWiFS uncertainties, the spectral measurements all fall within the individual channel k = 1 uncertainties as well. Since there is no statistically meaningful way to distinguish any spectral dependence within the measurement uncertainties, the means of the spectral measurements are used for the phase correction, as shown in Figure 6c; the standard deviations of the spectral means are given by the error bars in the figure and are highlighted separately in Figure 6d.

Comparing the SeaWiFS and PLEIADES phase corrections to the ROLO Model, both set to 0 at a phase angle of +7°, there is a significant offset but a similar dependence in the phase corrections from the two instruments, Figure 7. To develop a generalized phase correction to the output of the ROLO Model, the SeaWiFS data set is overlapped with the PLEIADES data set. A bias of +1.5 % was applied to the PLEIADES data at negative phase angles and a bias of 0.75 % was applied to the PLEIADES data at positive phase angles to overlap the two data sets. The resultant phase correction of the combined data sets are shown in Figure 8. To fit the unified data set, only SeaWiFS data are used between -30° and +30° phase; both data sets are used for phase angles less than -30° or greater than +30°. A cubic polynomial was used to fit the data set; results are shown in Figure 9.

3.3 Libration correction to the output of the ROLO Model

The effects of libration on the lunar irradiance are a strong function of phase angle, being a smaller effect at smaller phase angles [5], and are a function of wavelength [33]. Recently Eplee *et al*. re-examined the full SeaWiFS data set and applied a mean residual libration correction to the output of the ROLO Model of 0.18 % for low lunar phase angles [34]. The standard deviation of the libration correction over the SeaWiFS bands was 0.01 %. While the libration correction is wavelength dependent, spectral differences in the magnitude of the libration correction are within the k=1 measurement uncertainty. The libration correction is plotted as a function of band-center wavelength in Figure 10.

SNPP VIIRS measured the lunar irradiance 8 to 9 times per year, roughly at monthly intervals, at a nominal phase angle of -51° [33, 35]. The magnitude of the libration correction to the output of the ROLO Model for SNPP VIIRS at this phase angle is shown in Figure 10 (Table 6 in Ref. [33]). The error bars in the high phase libration correction to the output of the ROLO Model, given as the mean residuals in the lunar time series, are from Table 8 in Ref. [33]. The uncertainties in the libration correction are given in Figure 11. Libration corrections are only given for low and high phase angles. For other phase angles, we assume a rectangular probability distributions with limits being (SeaWiFS correction minus uncertainty) to (VIIRS correction plus uncertainty). The resultant uncertainties are shown in the figure as well.

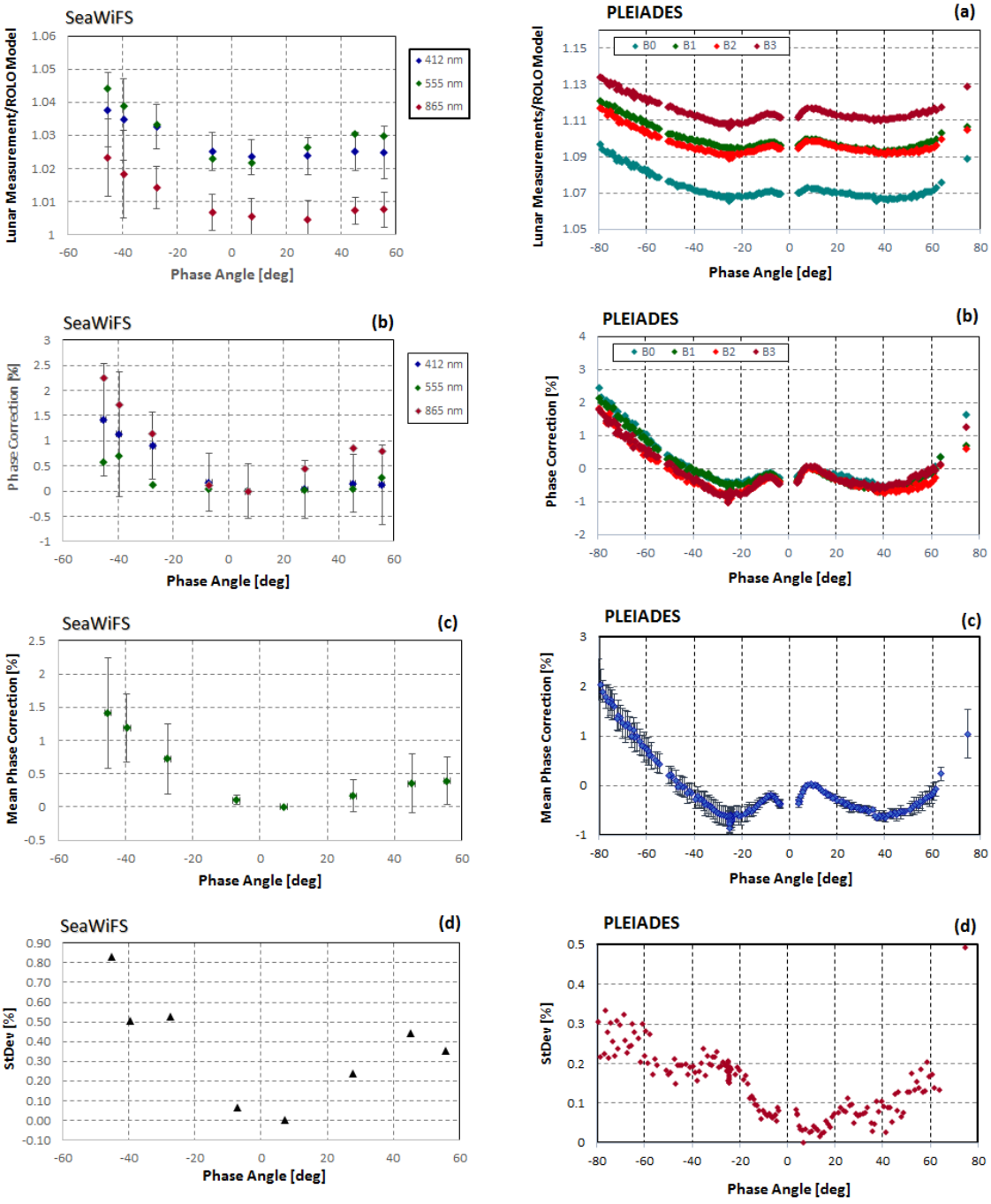


Figure 6. (a) Ratio between the lunar measurement and the output of the ROLO Model as a function of phase angle and wavelength; (b) Phase correction to the output of the ROLO model with all data set equal to 0 at a phase angle of +7°; (c) mean phase correction, averaged over wavelength. Error bars correspond to the standard deviation of the mean phase correction; (d) the standard deviation of the mean phase correction. (left) SeaWiFS and (right) PLEIADES.

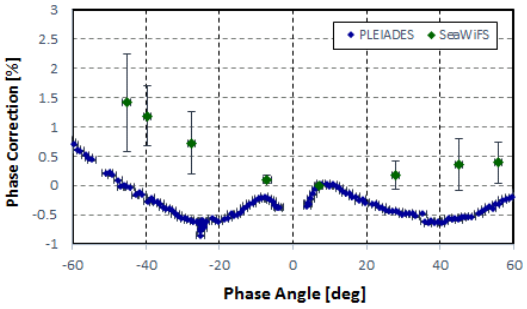


Figure 7. Correction to the phase dependence of the output of the ROLO Model for SeaWiFS and PLEIADES.

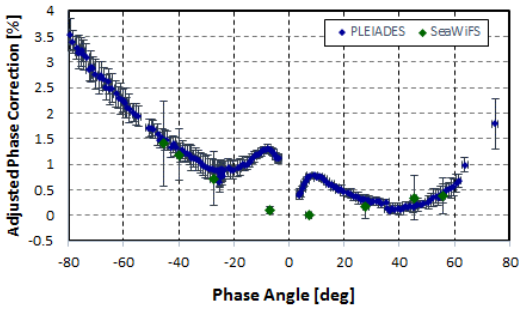


Figure 8. PLEIADES phase correction to the output of the ROLO Model adjusted to overlap with the SeaWiFS phase correction.

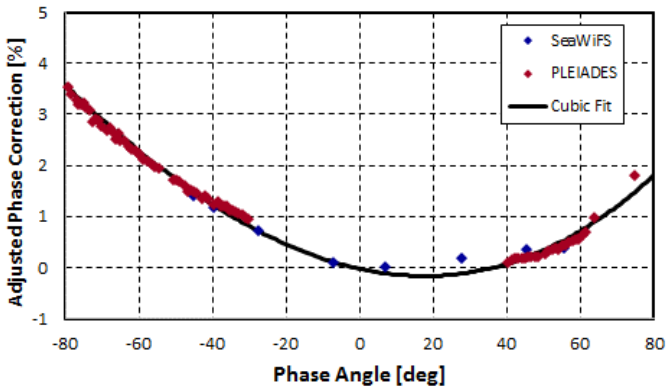


Figure 9. Cubic fit to the merged PLEIADES/SeaWiFS phase correction data sets. PLEIADES data were not used for phase angles between +40° and -30°.

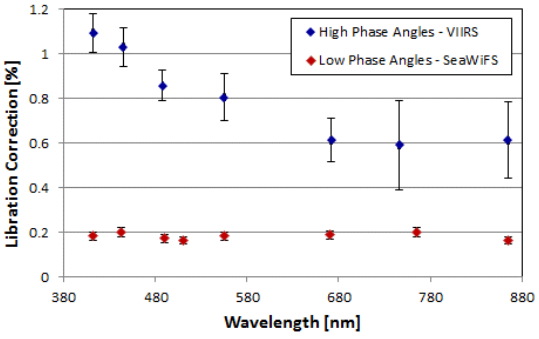


Figure 10. Magnitudes of libration corrections to the output of the ROLO model for high phase angles (VIIRS) and low phase angles (SeaWiFS).

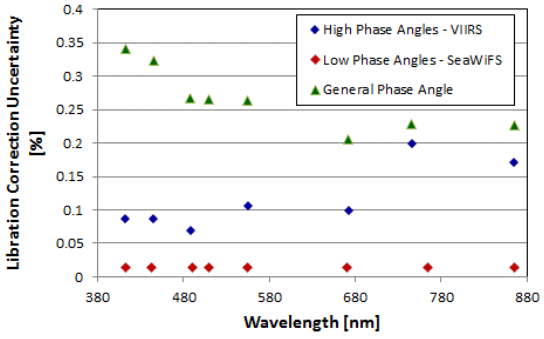


Figure 11. Uncertainty in the libration correction to the output of the ROLO Model for general phase angles, high phase angles, and low phase angles.

Brown, Steven; Eplee, Robert; Xiong, Xiaoxiong.
"Si-traceable top-of-the-atmosphere lunar irradiance: Potential tie-points to the output of the ROLO Model."
Paper presented at SPIE, San Diego, CA, United States. August 6, 2017 - August 11, 2017.

3.4 Uncertainty budget for the empirically corrected output from the ROLO Model

There are 3 components considered in the uncertainty budget: the uncertainty in the TOA lunar irradiance, the uncertainty in the phase correction to the output of the ROLO Model and the uncertainty in the libration correction to the output of the ROLO Model. While the corrections are a function of phase angle, uncertainty budgets are presented for low phase angles (Table 2) and high phase angles (Table 3).

There are three components to the phase correction uncertainty: a component for the spectral dependence of the phase correction, a component for the adjustment of the relative magnitudes of the phase correction between SeaWiFS and PLEIADES, and the uncertainty in the cubic fit to the phase correction of the blended SeaWiFS and PLEIADES data sets. The uncertainty in the spectral dependence to the phase correction is 0.1 % for low phase angles, using SeaWiFS data, and 0.2 % for high phase angles, using VIIRS data, line 2.1 in Tables 2 and 3.

For the purposes of these simplified uncertainty budgets, we take the uncertainty associated with merging the PLEIADES phase correction with the SeaWiFS phase correction to have a rectangular probability distribution, with limits of 0 % and 1.5 %, the adjustment to the PLEIADES data set for negative phase angles. The resultant uncertainty, defined to be the limits of the distribution divided by the $\sqrt{12}$,0.4 %, is listed in line 2.2 in Tables 2 and 3.

The cubic fit to the adjusted phase correction as a function of phase angle had a correlation coefficient of 0.993. The fit uncertainty was calculated using Eq. 4:

$$StDev = \sqrt{\frac{\Sigma\left(y_i - \overline{y}_i\right)}{n-1}} \ , \tag{4}$$

where y_i are the fit residuals, $\overline{y}_i$ is the mean of the residuals and n is equal to the number of data points. The fit uncertainty, 0.73 %, is given in line 2.3 of Tables 2 and 3.

The uncertainty in libration corrections to the output of the ROLO Model for low phase angles is given for SeaWiFS measurements [34] and the uncertainty in the libration correction for high phase angle was determined using VIIRS data [33]. The uncertainties are listed in line 3 of Tables 2 and 3.

The last lines in the tables give the combined standard uncertainties in the corrections to the output of the ROLO Model using an absolute tie point at +7° phase.

Table 2. Uncertainty budget for the corrected output from the ROLO Model for low phase angles.

Low Phase Angle (+7°)		Wavelength [nm]							
Element	Description	412	443	490	510	555	670	765	865
1	TOA Lunar Irradiance	2.68	0.93	0.62	0.60	0.48	0.40	0.38	0.37
2	Phase Correction	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
2.1	Spectral Dependence	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
2.2	SeaWiFS/PLEIADES adjustment	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
2.3	Cubic Fit	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73
3	Libration Correction	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
	Combined Standard Uncertainty [%]	2.81	1.26	1.04	1.04	0.97	0.93	0.94	0.93

Table 3. Uncertainty budget for the corrected output from the ROLO Model for high phase angles.

High Phase Angle (-51°)		Wavelength [nm]							
Element	Description	412	443	490	510	555	670	765	865
1	TOA Lunar Irradiance	2.68	0.93	0.62	0.60	0.48	0.40	0.38	0.37
2	Phase Correction	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
2.1	Spectral Dependence	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
2.2	SeaWiFS/PLEIADES adjustment	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
2.3	Cubic Fit	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73
3	Libration Correction	0.09	0.09	0.07	0.09	0.11	0.10	0.20	0.17
	Combined Standard Uncertainty [%]	2.81	1.27	1.06	1.05	0.99	0.95	0.96	0.95

4. DISCUSSION

Section 4.1 discusses the path forward toward further reductions in the uncertainty budget through examination of the current uncertainty budget and identification of the largest components in that budget. An example uncertainty budget is given showing the potential uncertainty in lunar irradiance measurements is given. In Section 4.2, an example of an unexpected impact of the corrected output from the ROLO Model is illustrated by a comparison of two laboratory calibration coefficients, one in 1993 by SBRS and the second in 1997 by NIST [36], and the SeaWiFS vicarious calibration-based gain coefficients using the Marine Optical Buoy (MOBY) data set [37].

4.1 TOA lunar irradiance path forward

Examination of the uncertainty budgets provides direction for future work to further reduce the uncertainty in measurements of the lunar irradiance and, as a consequence, to further reduce the uncertainty in the revised output from the ROLO Model. The uncertainty in the measured absolute TOA lunar irradiance and the fit to the phase correction dominate the uncertainty budget in Table 2. The uncertainty in the absolute TOA lunar irradiance is dominated by the uncertainty in the telescope calibration which is, in turn, dominated by the uncertainty in the irradiance responsivity of the reference spectrograph. Consequently, reducing the uncertainty in the reference spectrograph calibration will directly result in a reduction in the uncertainty in the absolute lunar irradiance. Using NIST's Facility for Spectral Irradiance and Radiance responsivity using Uniform Sources (SIRCUS), the uncertainty in the spectrograph calibration can be reduced to approximately 0.15 % [38]. Work at Mt. Hopkins has established the stability of the telescope responsivity to be 0.2 % over the course of the evening. Work by Eplee *et al*. has established the precision (residual Type A uncertainty) in the phase and libration corrections to be on the order of 0.1 %. Neglecting the uncertainty in the atmospheric correction and taking the Root-Sum-Square of the 4 components, the combined standard uncertainty in the TOA lunar irradiance could potentially be reduced to approximately 0.3 %.

4.2 Impact on Laboratory calibration coefficients

SeaWiFS was calibrated in the laboratory twice, once in 1993 by Raytheon Santa Barbara Remote Sensing (SBRS) and once in 1997 by NIST. While approaches to traceability to the SI differed, both calibrations used lamp-illuminated integrating spheres (LI-IS) as sources of spectral radiance [36]. No uncertainty budget or estimate was disseminated along with the 1993 calibration. A radiometric accuracy of 5 % (k=1) was one of SeaWiFS' performance specifications; this value is used as a general estimate of the uncertainty in the 1993 calibration. In 1997, during the NIST calibration of SeaWiFS, the SeaWiFS Transfer Radiometer (SXR) measured the radiance of the LI-IS before and after calibration of SeaWiFS using the LI-IS. The uncertainties in the 1997 pre-launch calibration of SeaWiFS were set at 4 % [39]. The two calibrations agreed within their combined uncertainties; coefficients of both the 1993 and the 1997 calibrations are given in Table 7 in Reference [36]. The NIST 1997 prelaunch calibration of SeaWiFS was used from the launch of the instrument through the R2007.0 reprocessing of the SeaWiFS database. In response to trends in the vicarious calibration with wavelength and the corresponding trends in lunar residuals using the ROLO Model, for the reprocessing of the SeaWiFS data set in 2009, R2009.0, the SBRS 1993 calibration was used [11]. The SBRS 1993 calibration has been used since then.

SeaWiFS, MODIS and VIIRS ocean color bands are all calibrated vicariously. Visible ocean color bands, for example SeaWiFS visible bands, bands 1-6, with band-center wavelengths ranging from 412.6 nm to 668.3 nm, are calibrated against MOBY-determined water-leaving radiances propagated to the top-of-the-atmosphere (TOA) using SeaWiFS-retrieved atmospheric correction parameters [37, 40]. The calibration approach for the NIR bands, SeaWiFS bands 7 (765 nm) and 8 (866.2 nm) assumes that the vicarious gain of the 865 nm band (Band 8 for SeaWiFS) is unity; Band 7 (SeaWiFS 765 nm band) is then calibrated vicariously relative to Band 8 [37, 40]. A similar approach is used for the MODIS and VIIRS instruments.

In Figure 12, three data sets are compared: the ratio between the SBRS93 laboratory calibration and the NIST97 laboratory calibration, the ratio between the output from the ROLO Model and the measurements by NIST at Mt. Hopkins, and the SeaWiFS vicarious calibration (Vc) coefficients derived using MOBY. Note that all 3 data sets show the same spectral trend. Error bars for the SeaWiFS Vc coefficients are from Table 2 in Franz *et al*. [37]; the uncertainty of the 865 nm Band is 0; an uncertainty of 0.01 was added to Band 8 for visualization purposes to bring the uncertainty in-line with the other bands.

The low uncertainty, TOA lunar irradiance measurements at Mt. Hopkins revealed a bias in the absolute output from the ROLO Model, Figure 5, that is very similar to the ratio between the two laboratory calibrations of SeaWiFS. Dividing the SeaWiFS Vc coefficients and the ROLO Model to NIST Hopkins ratio by the SBRS93 to NIST97 calibration coefficients ratio reverts back to use of the NIST97 calibration. Figure 13 shows the result for the SeaWiFS Vc coefficients and for the ratio between the output of the ROLO Model and the Mt. Hopkins measurements. There is no longer a clear spectral dependence to the ratios. All bands are within 2 %, with the exception of Band 8, where the Vc ratio to the ratio of the laboratory calibrations is off by 3.5 % and the ROLO to Hopkins ratio divided by the ratio of the laboratory calibrations is off by approximately 2.2 %. The results presented in Figs. 12 and 13 are suggestive and may lead to a reconsideration of the laboratory calibration coefficients used for SeaWiFS; specifically, returning to the NIST97 calibration.

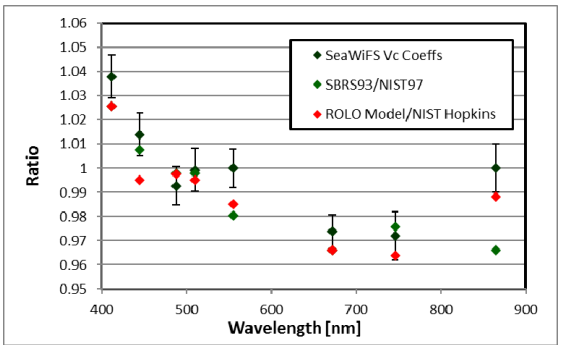


Figure 12. (dark diamonds) SeaWiFS vicarious calibration coefficients derived using the Marine Optical Buoy. (green diamonds) Ratio between the SBRS 1993 calibration of SeaWiFS and the 1997 NIST calibration of SeaWiFS. (Red diamonds) Ratio between the output from the ROLO model and the NIST measurements at the Whipple Observatory on Mt. Hopkins.

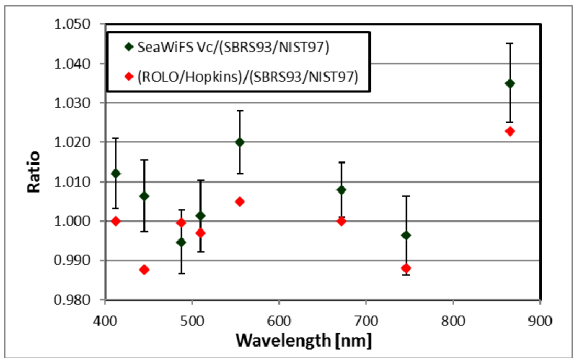


Figure 13. (dark diamonds) SeaWiFS Vicarious calibration gains and (red diamonds) Ratio of the output of the ROLO model to the NIST measurements at the Whipple Observatory on Mt. Hopkins ratio'd to the ratio between the SBRS 1993 and the NIST 1997 laboratory calibrations of SeaWiFS.

5. SUMMARY AND CONCLUSIONS

The GSICS Research Working Group, in collaboration with the CEOS Infrared and Visible Optical Sensors (IVOS) subgroup, held a Lunar Calibration Workshop in December 2014 [10]. One of the primary objectives of the Workshop was to provide the international community with a validated and SI-traceable rendition of the ROLO Model, version

3.11g (the version of the Model used in this work). A Workshop goal was to establish an absolute SI-traceable lunar irradiance scale for on-orbit satellite sensor lunar calibration with an uncertainty less than approximately 1 %.

This work has developed an approach for the establishment of an absolute TOA lunar irradiance model whereby the output of the ROLO Model at a phase angle of 7^o is tied to the absolute SI-traceable TOA lunar irradiance measured at Mt. Hopkins at a phase angle of 6.6^o. Combining the absolute lunar irradiance tie-points with phase corrections to the output of the ROLO Model using SeaWiFS and PLEIADES measurements and libration corrections based on SeaWiFS and SNPP VIIRS measurements, the resultant empirically corrected output from the ROLO Model has uncertainties of approximately 1 % through much of the silicon range.

Additional measurements of the lunar irradiance as a function of phase angle would help constrain the uncertainties in the phase correction to the output of the ROLO Model. To make additional low uncertainty measurements of the lunar irradiance as a function of phase angle, a project has been funded by NASA to make lunar irradiance measurements above 90 % of the atmosphere from an ER-2 aircraft [41]. A series of 2 week-long science missions are planned. The lunar phase changes by approximately 12^o/night; over 5 nights, the lunar irradiance could be measured over a 60^o range of phase angles. Concurrently, NIST researchers are in the process of developing a Lunar Observatory at the NOAA site on Mauna Loa (MLO). An absolute, SI-traceable, revised lunar model is the expected result from the multi-year time series of lunar measurements at MLO.

While ROLO measured the lunar radiance/irradiance, the ROLO Model is a reflectance model requiring an input solar spectral irradiance to use it as an on-orbit calibration target. Due to the current uncertainties in the solar spectral irradiance [42, 43], use of the Moon as an absolute exo-atmospheric reference calibration target for Earth remote sensing sensors may require additional, low uncertainty measurements of the solar spectral irradiance if it is to remain a reflectance model [44, 45]. Looking to the future, extending the lunar spectral measurement capability to the 1000 nm to 2500 nm spectral range and making lunar phase angles from approximately $+10^o$ to $+60^o$ and from -10^o to -60^o in roughly 10^o steps to establish the phase correction to the ROLO Model in the near-infrared spectral region would be beneficial to the Earth remote sensing community. The success of the ROLO Model is illustrated by the number of Earth-observing instruments, both U.S. and international, looking at the Moon and utilizing the Model to trend on-orbit temporal degradation in sensor responsivity. Development of an empirically corrected, absolute, SI-traceable revised ROLO Model with sub-1 % uncertainties should serve to expand the number of applications of a lunar irradiance/reflectance model relevant to Earth remote sensing sensors.

ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank Claire C. Cramer, United States Department of Energy, for providing the Mt. Hopkins lunar irradiances and uncertainty budgets and Aime Meygret, Centre National d'Etudes Spatial, France, for providing PLEIADES data.

REFERENCES

1. Kieffer, H.H., *Photometric stability of the lunar surface.* Icarus, 1997. **130**: p. 323-327.
2. Kieffer, H.H. and T.C. Stone, *The spectral irradiance of the moon.* Astron. J., 2005. **129**: p. 2887-2901.
3. Anderson, J.M., H.H. Kieffer, and K. Becker, *Modeling the brightness of the Moon over 350-2500 nm for spacecraft calibration.* Proc. SPIE, 2001. **4169**: p. 248-259.
4. Stone, T.C. and H.H. Kieffer, *Absolute irradiance of the Moon for on-orbit calibration.* Proc. SPIE, 2002. **4814**: p. 211-221.
5. Kieffer, H.H. and R.L. Wildbey, *Establishing the Moon as a spectral radiance standard.* J. Atmos. Oceanic Technol., 1996. **13**: p. 360-375.

6. Stone, T.C., *Stellar calibration of the ROLO lunar radiometric reference.* Proc. SPIE, 2010. **7807**: p. 78010T.
7. Wehrli, C., *Extraterrestrial solar spectrum*. WRC Publication. Vol. 615. 1985, Davos Dorf, Switzerland: Physikalisch-Meteorologisches Observatorium/World Radiation Center (PMO/WRC). 7 pp.
8. Stone, T.C. and H.H. Kieffer, *Assessment of uncertainty in ROLO Lunar Irradiance for on-orbit calibration.* Proc. SPIE, 2004. **5542**: p. 300-310.
9. http://oceancolor.gsfc.nasa.gov/cms/. The OBPG serves as a Distributed Active Archive Center for satellite ocean biology data produced or collected under NASA's Earth Observing System Data and Information System.
10. Wagner, S.C., et al., *A summary of the joint GSICS-CEOS/IVOS lunar calibration workshop: working towards intercalibration using the Moon as a transfer target.* Proc. SPIE, 2015. **9639**: p. 96390Z-1.
11. Eplee, J., R. E., et al., *On-orbit Calibration of SeaWiFS.* Appl. Opt., 2012. **51**: p. 8702-8730.
12. Eplee, J., R. E., et al., *Uncertainty assessment of the SeaWiFS on-orbit calibration.* Proc. SPIE, 2011. **8153**: p. 8153OB-1.
13. Hayes, D.S., *Stellar absolure fluxes and energy distributuions from 0.32 to 4.0 microns*, in *Calibration of Fundamental Stellar Quantities: Proceedings of the IAU Symposium No. 111*. 1985, D. Reidel Publishing Co.: Dordrecht. p. 225-252.
14. Castelli, F. and R.L. Kurucz, *Model atmospheres for Vega.* Astron. Astrophys., 1994. **281**: p. 817-832.
15. Peterson, D.M., et al., *Vega is a rapidly rotating star.* Nature, 2006. **440**: p. 896-899.
16. CEOS <http://calvalportal.ceos.org/home>
17. GSICS <http://gsics.wmo.int/>
18. Ohring, G., et al., *Satellite instrument calibration for measuring global climate change*. 2004, NISTIR7047: Gaithersburg, MD.
19. Cooksey, C. and R. Datla, *Workshop on bridging satellite climate data gaps.* J. Res. Natl. Inst. Stand. Technol., 2011. **116**: p. 505-512.
20. Zibordi, G., et al., *System vicarious calibration for ocean color climate change applications: Requirements for in situ data.* Remote Sensing of Environment, 2015. **159**: p. 361-369.
21. Thome, K.J., T. Gubbels, and R.A. Barnes, *Preliminary error budget for the reflected solar instrument for the Climate Absolute Radiance and Refractivity Observatory.* Proc. SPIE, 2011. **8153**: p. 81530R-1.
22. Thome, K.J., et al., *Test plan for a calibration demonstration sysetm for the reflected solar instrument for the Climate Absolute Radiance and Refractivity Obseratory.* Proc. SPIE, 2012. **8516**: p. 851602.
23. Cramer, C.E., et al., *Precise measurement of lunar spectral irradiance at visible wavelengths.* J. Res. Nat'l. Inst. Stds. Technol., 2013. **118**: p. 396-402.
24. Barnes, R.A., et al., *SeaWiFS Prelaunch Radiometric Calibration and Spectral Characterization*. NASA Tech. Memo. 104566, ed. S.B. Hooker, E.R. Firestone, and J.G. Acker. Vol. 23. 1994, Greenbelt, MD: NASA Goddard Space Flight Center. 55.
25. Lacherade, S., et al. *PLEIADES absolute calibration: inflight calibration sites and methodology*. in *XXII ISPRS Congress*. 2012.
26. Oudrari, H., et al., *Prelaunch radiometric characterization and calibration of the S_NPP VIIRS sensor.* IEEE Transactions on Geoscience and Remote Sensing, 2015. **55**(4): p. 2195-2210.
27. Cramer, C.E., et al., *A novel apparatus to measure reflected sunlight from the Moon.* Proc. SPIE, 2013. **8867**: p. 8867OW.

28. Yoon, H.W. and C.E. Gibson, *Spectral irradiance calibrations*. Natl. Inst. Stand. Technol. Spec. Publ. Vol. SP250-43. 2011, Washington, DC: U.S. Bovernment Printing Office. 123.
29. Zong, Y., et al., *Simple stray light correction method for array spectroradiometers.* Appl. Opt., 2006. **45**(6): p. 1111-1119.
30. Kieffer, H.H., et al., *On-orbit radiometric calibration over time and between spacecraft using the Moon.* Proc. SPIE, 2003. **4881**: p. 287-297.
31. Lacherade, S., et al., *POLO: a unique dataset to derive the phase angle dependence of the Moon irradiance.* Proc. SPIE, 2014. **9241**.
32. Eplee, J., R. E., et al., *Cross calibration of SeaWiFS and MODIS using on-orbit observations of the Moon.* Appl. Opt., 2011. **50**(2): p. 120-133.
33. Eplee, J., R. E., et al., *On-orbit calibration of the Suomi National Polar-orbiting Partnership Visible Infrared imaging Radiometer Suite for ocean color applications.* Appl. Opt., 2015. **54**(8): p. 1984-2006.
34. Eplee, J., R. E., F.S. Patt, and G. Meister, *Geometric effects in SeaWiFS lunar observations.* Proc. SPIE, 2015. **9607**: p. 960704.
35. Eplee, J., R. E., et al., *Updates to the on-orbit calibration of SNPP VIIRS for ocean color applications.* Proc. SPIE, 2015. **9607**.
36. Barnes, R.A. and E.F. Zalewski, *Reflectance-based calibration of SeaWiFS. II. Conversion to Radiance.* Appl. Opt. **42**: p. 1648-1660.
37. Franz, B.A., et al., *Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry.* Appl. Opt., 2007. **46**: p. 5068-5082.
38. Brown, S.W., G.P. Eppeldauer, and K.R. Lykke, *Facility for spectral irradiance and radiance responsivity calibrations using uniform sources.* Appl. Opt., 2006. **45**(32): p. 8218-8237.
39. Johnson, B.C., et al., *Volume 4, The 1997 Prelaunch Radiometric Calibration of SeaWiFS*, in *SeaWiFS Postlaunch Technical Report Series*, S.B. Hooker and E.R. Firestone, Editors. 1999, NASA/TM-1999-206892.
40. Eplee, J., R. E., et al., *Comparison of SeaWiFS on-orbit lunar and vicarious calibrations.* Proc. SPIE, 2006. **6296**.
41. Turpie, K.R., Principal Investigator, *Development of a highly accurate lunar spectral irradiance measurement capabiliy (16-AITT16-0002)*. 2017.
42. Harder, J.W., et al., *The SORCE SIM solar spectrum: comparison with recent observations.* Solar Phys, 2010. **263**: p. 3-24.
43. Pagaran, J., et al., *Intercomparison of SCIAMACHY and SIM vis-IR irradiance over several solar rotational timescales.* A&A, 2011. **528**.
44. Thome, K.J., S.F. Biggar, and P.N. Slater, *Effects of assumed solar spectral irradiacne on intercomparisons of Earth-observing sensors.* Proc. SPIE, 2001. **4540**: p. 260-268.
45. Shanmugam, P. and Y.H. Ahn, *Reference solar irradiance spectra and consequences of their disparities in remote sensing of the ocean colour.* Ann. Geophys., 2007. **25**: p. 1235-1252.

H02: Low-Dimensional Nanoscale Electronic and Photonic Devices 10

Electronic and Chemical Structure of 2D materials

Sujitra Pookpanratana, EPD, NIST

The drive to produce smaller, faster, lower power electronic components for computing is pushing the semiconductor industry to nanometer-scale device structures. Atomically thin two-dimensional (2D) materials such as transitional metal dichalcogenides (TMDs) and graphene are a promising class of materials for nanoelectronics with unique optical and electronic properties. The properties of mono- to few-layer TMDs are highly sensitive and reactive to their local environment (ambient conditions and adsorbates) and the interfaces within a fabricated device structure (gate dielectric, contacts), which can all impact their electronic properties and extrinsic device characteristics. Here, we will present the electronic and chemical structure of doped graphene on silicon carbide and polymorphs of molybdenum telluride.

Large area epitaxial graphene (EG) grown on silicon carbide (SiC) provides a feasible route to scalable production of graphene-containing electronic components. EG on SiC display strong n-type (electron) doping due to the interaction of the zero-layer graphene [1] that is directly bonded to SiC. Nitric acid has been shown to be an effective p-doping agent and produce low carrier density of EG [2]. Using laboratory-based ultraviolet (UPS) and x-ray photoelectron (XPS) spectroscopies, we investigate the surface chemical composition and electronic structure of nitric acid exposed EG on SiC. We observe the presence of nitrogen on the surface which provides direct evidence on the role of nitric acid as extrinsic dopant of EG on SiC. The UPS results indicate that the EG surface is that of a semi-metal after exposure to nitric acid. Technological commercialization of graphene for electronic components requires that their electronic properties be engineered for specific device functionality such as the charge carrier density.

With the vast elemental phase space of TMDs, TMDs have a range of electronic properties giving rise to metallic, semiconducting, or insulating properties. In addition, some TMDs exhibit polymorphism and different packing geometry or phase can distinctly impact the electronic structure and it will influence optical and transport properties. Molybdenum telluride ($MoTe_2$) can exist in three phases: 2H, 1T, and 1T'. The 2H phase is semiconducting, while the other phases are metallic or semi-metallic. For the bulk 2H $MoTe_2$ crystal, the surface is relatively stable and with little oxidation based on our XPS and UPS results. This is not the case for the 1T' phase of crystalline $MoTe_2$, where the molybdenum and tellurium atoms at the surface are significantly oxidized. The 1T' phase can be cleaved at the surface, and we can observe the semi-metallic properties by angle-resolved photoemission. The cleaved 2H $MoTe_2$ surface exhibits dispersion of the valence bands and a valence band maximum consistent with being an n-type semiconductor. $MoTe_2$ is appealing for manufacturing of nanoelectronic components due to tunability in electronic properties in one material system. Continued innovation requires novel nanoelectronic materials and devices, coupled with a thorough understanding of the electronic structure which determines device functionality.

[1] P. Mallet, F. Varchon, C. Naud, L. Magaud, C. Berger, and J.-Y. Veuillen, Phys. Rev. B 76, 041403 (2007).

Pookpanratana, Sujitra.
"Electronic and Chemical Structure of 2D materials."
Paper presented at 232nd Electrochemical Society Meeting, National Harbor, MD, United States. October 1, 2017 - October 5, 2017.

[2] Y. Yan, L.-I. Huang, Y. Fukuyama, F.-H. Liu, M. A. Real, P. Barbara, C. -T. Liang, D. B. Newell, and R. E. Elmquist, Small 11, 90 – 95 (2015).

Pookpanratana, Sujitra.
"Electronic and Chemical Structure of 2D materials."
Paper presented at 232nd Electrochemical Society Meeting, National Harbor, MD, United States. October 1, 2017 - October 5, 2017.

METHOD TO DETERMINE THE CENTER OF CONTRAST TARGETS FROM TERRESTRIAL LASER SCANNER DATA

Prem Rachakonda*, Bala Muralikrishnan*, Daniel Sawyer*, Ling Wang**

*Dimensional Metrology Group,
Engineering Physics Division
National Institute of Standards and Technology,
Gaithersburg, MD, USA

**Department of Automation, School of Mechanical and Electrical Engineering,
China Jiliang University, Hangzhou,
Zhejiang Province, 310018, P. R. China

1 INTRODUCTION

Terrestrial laser scanners (TLSs) are instruments that can measure 3D coordinates of objects at high speed using a laser, resulting in high density 3D point cloud data. Contrast targets are some of the most common targets used with TLSs to establish a scale or register multiple scan datasets. These targets are also known as checkerboard targets or signalized targets and their design provides a way to calculate their geometric center by using intensity data along with the dimensional data. Large contrast targets are needed when scanning at longer distances; however, unlike other geometric targets such as spheres, fabrication of large contrast targets is relatively inexpensive. Even though contrast targets are used with TLSs, the algorithms to calculate their centers are proprietary and/or work only with proprietary data formats.

In this context, this paper provides a novel method that was developed at the National Institute of Standards and Technology (NIST) to calculate the derived point or the center of a contrast target (henceforth termed as CCT) and compares its performance with commercial software, in various scan conditions.

2 PROCEDURE TO OBTAIN THE CENTER OF THE CONTRAST TARGET

Calculating the center of a contrast target has not been studied extensively in open literature. This paper describes one method to obtain the center of a contrast target which involves multiple steps, where each successive step attempts to refine the center of the target obtained in the previous step. For the target depicted in Figure 1, the center is the location where the two black squares meet.

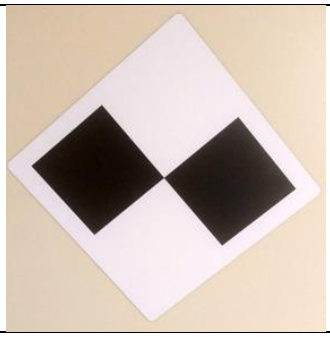


Figure 1: Picture of a contrast target

Most software that are provided by TLS manufacturers output intensity or color data along with dimensional data for their scans. For this procedure, the dimensional data was exported along with intensity data (XYZI format). Some TLSs provide dimensional data along with color intensity (XYZRGB). Here, X, Y, Z correspond to the dimensional data, I is the intensity data and R, G, B correspond to red, green, blue channel intensity data. If the data obtained is in the form of XYZRGB, the RGB data needs to be converted to intensity values using a weighted sum using the formula $I = 0.299 \times R + 0.587 \times G + 0.114 \times B$ [1]. Subsequently, the XYZI formatted data can be processed using the following steps to obtain the center of a contrast target.

1. Extract the data corresponding to the region of interest of the contrast target within the scene.
2. Calculate an approximate 3D location of the center of the target using 2D imaging methods.

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel; Wang, Ling.
"Method to Determine the Center of Contrast Targets from Terrestrial Laser Scanner Data."
Paper presented at 32nd ASPE Annual Meeting, Charlotte, NC, Charlotte, NC, United States. October 29, 2017 - November 3, 2017.

3. Obtain a refined location of the center of the target using dimensional and intensity data.

2.1 Extract targets' regions of interest

The first step in this process was to extract the regions of interest from a scan. This step lowers the complexity and number of computations needed to find the 2D center of the target. Figure 2 is the extracted intensity image of a scan which shows multiple checkboard targets placed around a room in different orientations. Data corresponding to individual targets need to be cropped/separated. This process needs to be performed for both the dimensional and the intensity data.

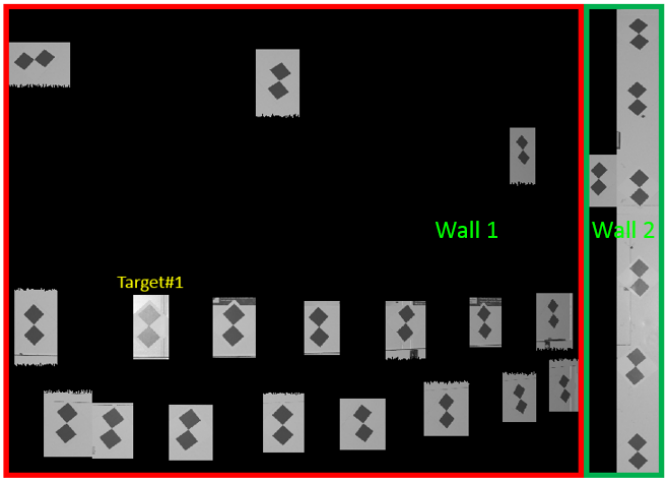


Figure 2: Intensity image from a TLS showing multiple targets mounted on two walls

This cropping of the data (XYZI) can be done either manually or automatically. The manual method may use any point cloud manipulating software for cropping, whereas the automatic methods may use a template matching method. For this work, all the regions of interest were cropped manually once for the first dataset and the same cropping regions were used to crop the remaining datasets. All subsequent processing described next was performed using automated methods.

2.2 Calculate the target's approximate 3D center using 2D imaging methods

To obtain the center of the target, the intensity information can be used to generate a 2D image of the target. Subsequently, this image is processed using edge detection methods along with the Hough transform [2]. These methods aid in detecting the intersecting lines in a checkerboard-like pattern [3]. The Hough transform is a mature algorithm and is built into software like MATLAB* which was used to process this data. As is usual with many image processing algorithms some level of adjustments is required when the nature of the image changes (low contrast, high noise etc.). The steps to extract an approximate target center (CCT) are as follows:

i. Each cropped XYZI dataset was processed and an image was generated to show a single contrast target as shown in *Figure 3*a.

ii. The image generated by the previous step was cropped further using a circular mask to display the central region where the two black squares touch each other. Care must be taken to avoid any other regions with intersecting lines/corners as shown in *Figure 3*b. The masking was done automatically using a predefined mask position and size. Using a square mask instead of a circular mask would lead to detection of additional lines that are at the edges of the square mask.

* Disclaimer: Commercial equipment and materials may be identified in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel; Wang, Ling.
"Method to Determine the Center of Contrast Targets from Terrestrial Laser Scanner Data."
Paper presented at 32nd ASPE Annual Meeting, Charlotte, NC, Charlotte, NC, United States. October 29, 2017 - November 3, 2017.

iii. Lines/edges were emphasized in masked image using a "Canny edge" algorithm.

iv. A Hough transform was performed on the image from the previous step and the intersecting lines are obtained. It is possible that there may be more than two lines due to the nature of the image (unclean targets, uneven target plane, shadows etc.). Care must be taken to apply appropriate filters to obtain only two intersecting lines at the center of the target.

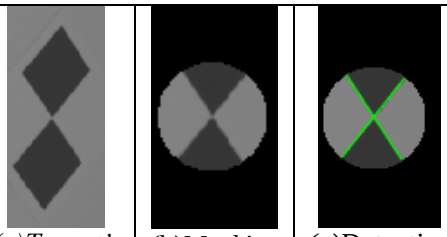
v. The two lines obtained in the previous step were intersected to obtain a 2D approximation of the center of the contrast target CCT_1 as shown in *Figure 3*c.

In the event of failure of the 2D method described above, a manual method may be used to pick the approximate 2D center from the image. It should be noted that CCT_1 was obtained from image data in units of pixels. Even though the pixels are integer numbers, CCT_1 may not have integer values.

Each pixel in the 2D image has a corresponding 3D coordinate and an intensity value. To obtain an approximate 3D center, the pixel center coordinates CCT_1 are rounded and an integer pixel location CCT_2 closest to the CCT_1 was obtained. This modified pixel location has a corresponding 3D coordinate X_3, Y_3, Z_3 and intensity value I_3. The 3D coordinate corresponding to CCT_2 was then $CCT_3 = (X_3, Y_3, Z_3)$.

2.3 Calculate a refined target center using dimensional and intensity data

One method to calculate the 3D center of the contrast target was described in the previous sub-section. When multiple scans of the same target were processed, the approximate 3D coordinates of the centers (CCT_3) were found to have poor repeatability ($1\sigma > 1$ mm), much larger than the expected repeatability of the system and the setup. To improve the calculation of target centers, a new method was conceived which uses intensity data along with dimensional data in non-radial directions. To perform this, first the 3D data in Cartesian coordinate system was converted to data in a spherical coordinate system of form (H, V, R), where H is the horizontal/azimuth angle, V is the vertical/elevation angle and R is the radial distance to each measured point.



(a)Target's image | *(b)Masking of image* | *(c)Detecting lines*

Figure 3: Use of Hough transform to detect the intersecting lines of a contrast target.

A new dataset was created that was comprised of the angles H, V and the intensity data I. Henceforth, this will be referred as the HVI domain in this paper. The radial data (R) was ignored at this point. To improve the results, the following modifications were performed on the data:

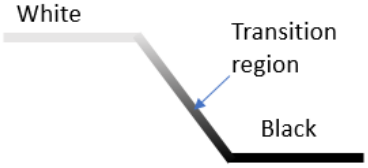


Figure 4: Illustration of the transition or fall-off regions of a contrast target

1. The density of data in the HVI domain was increased to 300 times the original point density by interpolation using a cubic polynomial. This was found to lower the uncertainty in calculating the intersection point. Although cubic polynomial interpolation has a possibility of wild swings in the interpolated points, the data was visually inspected after interpolation to ensure that such issues do not exist in the region of interest (fall-off region in Figure 4). Other

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel; Wang, Ling.
"Method to Determine the Center of Contrast Targets from Terrestrial Laser Scanner Data."
Paper presented at 32nd ASPE Annual Meeting, Charlotte, NC, Charlotte, NC, United States. October 29, 2017 - November 3, 2017.

interpolation techniques were also explored, however cubic interpolation was observed to perform adequately for most datasets.

2. In an ideal scan of a contrast target, the intensity data typically has a minimum value of 0 (black) and maximum value of 1 (white). However, most scans of contrast targets don't have such values for their black and white regions. For example, the intensity may range from 0.25 to 0.75. To lower the uncertainty of determining the CCT, intensity data was normalized/scaled from 0 to 1.

In the HVI domain, the intensity values corresponding to the intersecting lines in a 2D image have a sharp fall-off as illustrated in Figure 4. In this fall-off region, the intensity drastically changes from $I \approx 1$ to $I \approx 0$. These regions in the HVI domain are shown in Figure 5 where the colored regions in the middle are the high density interpolated points corresponding to the fall-off region. The following steps are then performed to obtain the intersection point:

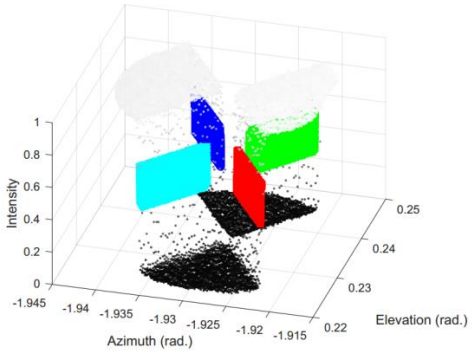


Figure 5: Transition regions of the target (colored) in the HVI domain that are in between the extreme intensity values.

a. The fall-off region in the HVI domain was obtained by discarding the surfaces corresponding to extreme intensities (black and white in Figure 5) that are over 0.5 times the standard deviation (0.5 σ) from the mean intensity value. This value of 0.5 σ was empirically determined and it ensured that only data belonging to the intersecting fall-off regions were obtained.

b. This fall-off region was then split into four parts after truncating the central cylindrical region close to the approximate center. This is performed using an automated method and this truncation enables the intersecting regions to be separated into four parts (as shown by the colored regions in Figure 5).

c. The intensity data from four datasets in the HVI domain was then discarded, keeping only H and V. This process essentially collapses the data to a single plane in HV domain, yielding datasets corresponding to four lines on the target.

d. In this step, a 2D intersection point in the HV domain was obtained. This can be obtained in one of three ways:

   i. The four line datasets can be intersected using a least-squares method in the HV domain to yield the center of the target (H_c, V_c).

   ii. The four line datasets can be grouped together into two datasets corresponding to two lines. These datasets can be least-squares fit to two lines, which can then be intersected to obtain (H_c, V_c).

   iii. Alternatively, before step#c, the four surfaces in the HVI domain (colored regions in Figure 5) can be fit to planes. These four plane equations can be solved using a least-squares method to perform a four-plane intersection. This method results in an intersection point (H_c,V_c,I_c) and thereby obtaining a center (H_c, V_c) in 2D.

   In general, it was found that 2D line intersection of two lines was more repeatable than the other two methods to yield an intersection point (H_c, V_c). Both the plane and line fitting routines were performed iteratively after excluding data points whose corresponding

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel; Wang, Ling.
"Method to Determine the Center of Contrast Targets from Terrestrial Laser Scanner Data."
Paper presented at 32nd ASPE Annual Meeting, Charlotte, NC, Charlotte, NC, United States. October 29, 2017 - November 3, 2017.

residuals exceeded three times the standard deviation of the residuals. The iterations were terminated when there were no more points to exclude.

e. The radial value of the center (R_c) was the radial value of CCT_3 in spherical coordinate system. It should be noted that CCT_3 was calculated using 2D imaging methods described in section 2.2. This value of R_c is an approximate value and will be improved in the subsequent steps.

f. The center (H_c, V_c, R_c) was then converted to a Cartesian coordinate system to obtain CCT_4 = (X_c, Y_c, Z_c).

g. As a final step, CCT_4 was projected onto the plane of the contrast target in the XYZ domain. This was performed by intersecting the line joining the origin and CCT_4 with the plane of the contrast target. This projected point CCT_5 = (X_p, Y_p, Z_p) was the final center of the contrast target. This projection method was required to ensure that the final center CCT_5 lies on the plane of the contrast target.

3 TEST METHODOLOGY

To understand the currently available methods and to compare them with the method described in this paper, three commercial software packages were used and repeatability studies were conducted. Two contrast targets were scanned 10 times and their CCTs and the standard deviations (1σ) of those CCTs were calculated using each software. Lower 1σ values of all the 3D coordinates (σx, σy, σz) indicates a more robust algorithm. This metric however considers only the precision of the centers but not their accuracy. It should be noted that there may be other software which may perform better, but were inaccessible to the authors at the time of writing this paper. One commercial software (Method#1) yielded lowest 1σ values (lower by an order of magnitude) and this software was used to compare the method described in this paper (Method#2).

Table 1: Comparison of the target center parameters

Target #	Method#1 (μm)				Method#2 (μm)				L (m)	M	D	θ°
	σH1	σV1	σR1	σD1	σH2	σV2	σR2	σD2				
1	48	76	40	53	75	96	34	63	4.4	1.2	1.2	2
2	37	68	29	51	35	76	26	53	4.5	1.0	1.0	11
3	48	71	33	53	47	65	28	47	4.5	0.9	0.9	6
4	54	69	30	53	50	70	27	52	4.5	0.9	1.0	6
5	57	80	27	59	50	80	24	57	4.7	0.9	1.0	15
6	48	72	36	47	46	69	30	46	4.7	0.9	1.0	16
7	35	67	28	37	35	64	27	41	4.7	1.0	1.1	19
8	42	85	36	62	41	82	34	62	4.8	1.0	1.0	22
9	16	86	28	55	12	81	25	53	4.9	0.9	1.0	15
10	8	64	24	34	14	63	23	34	5.0	1.2	1.0	18
11	60	87	32	72	57	86	28	69	5.0	0.9	1.0	26
12	39	111	21	72	46	118	25	76	5.1	1.1	1.1	5
13	55	123	29	87	55	121	24	84	5.1	0.9	1.0	6
14	38	94	33	67	40	94	30	65	5.2	1.0	1.0	31
15	18	108	22	70	23	107	21	71	5.2	1.1	1.0	5
16	64	124	29	87	66	120	28	85	5.2	1.0	1.0	5
17	13	120	22	75	19	119	19	75	5.4	1.1	1.0	5
18	56	102	34	75	56	100	29	74	5.4	0.9	1.0	34
19	34	129	17	78	20	123	19	76	5.6	0.9	1.0	4
20	24	108	39	71	23	115	30	70	5.7	0.9	1.0	23
21	48	109	34	75	47	107	36	76	5.7	1.0	1.0	39
22	66	115	30	82	61	109	29	80	6.2	1.0	1.0	44
23	25	111	30	69	23	116	26	73	6.3	1.0	1.1	44
24	51	122	31	81	50	119	24	81	6.4	0.9	1.0	47
25	65	120	30	88	63	118	34	85	6.6	1.0	1.0	47

3.1 Test setup

The test setup involved placing 25 contrast targets on two walls as depicted in Figure 2. These targets were square in shape, ≈ 225 mm wide, fabricated out of a flexible plastic material and have a magnetic backing for the purposes of mounting. Seven targets were mounted on a wall (green wall) on the right distributed vertically and the rest are mounted on a wall that was perpendicular (red wall), distributed horizontally. Even though these targets appear to be in the same plane in Figure 2 (planar view), they are in fact on two walls that are perpendicular. The TLS

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel; Wang, Ling.
"Method to Determine the Center of Contrast Targets from Terrestrial Laser Scanner Data."
Paper presented at 32nd ASPE Annual Meeting, Charlotte, NC, Charlotte, NC, United States. October 29, 2017 - November 3, 2017.

was placed at approximately 5 m from both the walls and the distance to each target (L) and the angles of incidence (θ) are listed in *Table 1*. The angle of incidence θ is the angle between the target's surface normal and the laser beam at the target's nominal center.

To perform an evaluation, these targets were scanned 10 times using a TLS. One such scan is depicted in Figure 1. Data acquired by this TLS was exported both to its own proprietary file format and to the XYZI format. Commercial software (Method#1) was used to process the data in this proprietary file format and the CCT of all the targets were obtained. This software did not have the capability to process the data in the XYZI format.

After processing the data, the standard deviation of the centers in spherical coordinates calculated by Method#1 was σ_1, and as calculated by Method#2 was σ_2. Here $\sigma_1 = (\sigma_{A1}, \sigma_{E1}, \sigma_{R1})$ and $\sigma_2 = (\sigma_{A2}, \sigma_{E2}, \sigma_{R2})$. To perform a comparison in the units of length, the standard deviations in azimuth and elevation were multiplied by the average radial distance value of the center. i.e., $\sigma_{H1} = R_1 \times \sigma_{A1}$, $\sigma_{V1} = R_1 \times \sigma_{E1}$, $\sigma_{H2} = R_2 \times \sigma_{A2}$, $\sigma_{V2} = R_2 \times \sigma_{E2}$. Here R_1 and R_2 were the average radial distances of the center as determined by Method#1 and Method#2 respectively. Two other parameters, σ_{D1} and σ_{D2} were calculated, which were the 1σ values of the distances of the centers about the mean value of the center calculated by Method#1 and Method#2 respectively. These are listed in *Table 1* along with a parameter $D = \sigma_{D2}/\sigma_{D1}$.

Ideally, the corresponding standard deviation values of both the methods (σ_1 and σ_2) should be equal, but they are not. To make the comparison simpler, a quality factor M given by equation 1 was introduced to compare the methods at each CCT and *Table 1* shows all the parameters calculated using both the methods using the 10 repeat measurements. This method of comparison is useful since Method#1 was found to be consistently producing centers with lower variation among the three commercial software packages that were evaluated.

$$M = \frac{M_{AZ} + M_{EL} + M_{RR}}{3} \qquad 1$$

where, $M_{AZ} = \frac{\sigma_{A2}}{\sigma_{A1}}$, $M_{EL} = \frac{\sigma_{E2}}{\sigma E_1}$ and $M_{RR} = \frac{\sigma_{R2}}{\sigma_{R1}}$.

If the standard deviation values from both the methods are identical, M = 1. If Method#1 performs better than Method#2 then M > 1 and *vice versa* if M < 1. There may be cases where M = 1 if Method#1 outperforms in one component and underperforms in another. This metric M gives a good estimate of the overall method performance and M_{AZ}, M_{EL}, M_{RR} reveal the performance of the method in individual spherical coordinate components.

3.2 Summary and discussion of the results

The centers of the contrast targets were calculated using a method involving multiple steps. The steps described in this paper are summarized below:

- CCT_1: 2D center using image processing methods
- CCT_2: 2D center with integer values of CCT_1
- CCT_3: 3D center corresponding to CCT_2
- CCT_4: 3D center using HVI domain method
- CCT_5: CCT_4 projected on the target's plane

Two notable trends affected the quality of Method#2. First, it was observed that the value of M

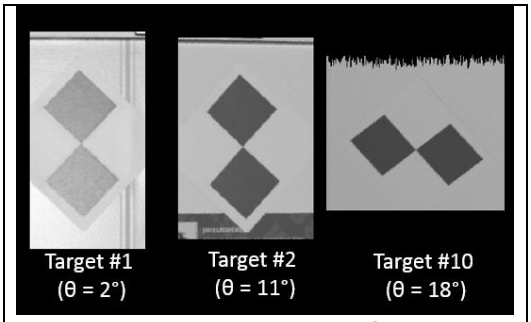


Figure 6: Intensity images of targets at various angles(θ) and orientations.

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel; Wang, Ling.
"Method to Determine the Center of Contrast Targets from Terrestrial Laser Scanner Data."
Paper presented at 32nd ASPE Annual Meeting, Charlotte, NC, Charlotte, NC, United States. October 29, 2017 - November 3, 2017.

was *not* significantly dependent on the angle of incidence (θ), but on the intensity variations in the black and the white regions of the target as shown in Figure 6. Targets that were placed at $\theta < 7°$ (shallow incidence angles) resulted in high intensity regions in the black part of the target. Second, target #10 also showed higher value of M as it was oriented differently compared to all the other 24 targets (see Figure 6). However, overall it can be observed from *Figure 7*b that Method#2 performs reasonably well in all the three components.

The distances between the centers obtained from Method#1 and Method#2 were calculated and the average distance between the centers at all the locations was ≈ 0.14 mm. There was no systematic bias that was observed in the azimuth or elevation coordinate differences of centers from both the methods, but there was a bias in the radial direction, an average of ≈ 0.12 mm. This bias could be a result of ensuring that the CCT is on the least-squares fitted plane of the target. It should be noted that the accuracy of either method cannot be ascertained using a single point measurement. Such a comparison would require test procedures involving calibrated lengths between two contrast targets in various orientations.

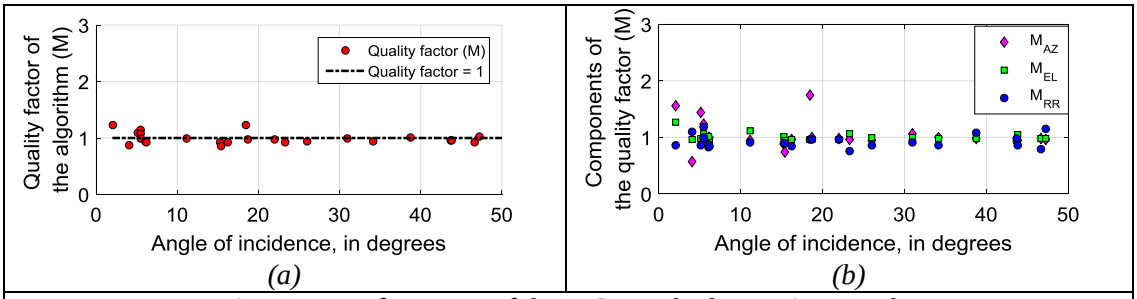


Figure 7: Performance of the NIST method at various angles

4 CONCLUSION

This paper presents a novel method developed at NIST to calculate the center of a contrast target and compares it with the results from other available software. It was observed that the NIST method performs reasonably well for most targets. More work is planned to ascertain the method's performance for various test cases (target orientations, data densities etc.). This is to ensure that the parameters used to deduce the target centers are more robust and applicable for scans with varying data quality.

5 REFERENCES

[1] ITU-R BT.601-71 - Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf (Retrieved 8/2/2017)

[2] Duda, R., Hart, P., "Use of Hough Transformation to Detect Lines and Curves in Pictures", Communications of the ACM, January 1972, Vol.15, No. 1

[3]http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/example.html (Retrieved 6/21/2017)

PRELIMINARY STUDY ON SPECTRAL CHARACTERISTICS FOR IDENTIFICATION OF SKIN COLOR UNDER CIRCULATORY DYSFUNCTION USING ARTIFICIAL SKIN SAMPLES

Akizuki, Y.[1], Ohno, Y.[2]

[1] Faculty of Human Development, University of Toyama, Toyama, Japan
[2] National Institute of Standards and Technology, Gaithersburg, MD USA
akizuki@edu.u-toyama.ac.jp

Abstract

To assist development of appropriate light sources to diagnose circulatory dysfunction conditions effectively based on the color appearance of patient's skin under emergency situation in disaster sites, a set of urethane skin model samples have been developed, which simulate visual characteristics of real human skins under different circulatory dysfunction conditions (shocked, healthy, and congested) for young female/male and elderly female/male. Their characteristics have been evaluated by calculations of samples' color differences under various light spectra compared to those of real skins. Visual evaluation experiments have been conducted using these samples to evaluate distinction between different circulatory dysfunction conditions under the optimized light-emitting diode (LED) light spectra used in our previous study. The developed urethane samples were found better in closeness of colors to real skin than previously-used paper color charts, but it needs further improvements in matching spectral reflectance characteristics. The experimental method directly comparing perceived color differences of the samples under the LED lights and reference lights (broadband spectra) was found useful for such evaluation.

Keywords: skin color, spectral distribution, circulatory dysfunction, emergency medical care

1 Introduction

In a disaster situation, in confined spaces after earthquake, typhoon, or tsunami disasters, victims can become a crush syndrome and need an immediate treatment at the rescue site. Under such a situation, it is necessary to judge correctly whether the victims rescued from debris have fallen into circulatory dysfunction and to apply appropriate medical treatment, in order to increase the survival rate. Unfortunately, it is difficult to judge their physical conditions by visual observation at the disaster site often due to poor performance of their lighting equipment. Also in a non-disaster medical environment such as in hospitals, where conventional lighting (such as by fluorescent and incandescent lamps) is being replaced by LED lighting, development of LED light sources enabling the medical staff to judge patients' conditions more accurately with enhanced color rendition performance than conventional light sources is desired.

In an effort to assist development of such light sources, a database of spectral reflectance data of circulatory dysfunctional skin colors was developed, taking age and gender into consideration, by measuring Japanese subjects' skins (at back of hands) which were artificially shocked or congested by controlling blood flow of an upper arms (Akizuki et al. 2013). Also, we conducted simulations to determine the spectral power distributions (SPD) of a theoretical LED source (three narrow bands) that maximize the color differences between healthy skin and shocked/congested skin for different age/gender groups, and conducted preliminary visual evaluation of the optimized spectra (chosen closest to the theoretical LED spectra as above) using NIST Spectrally Tunable Lighting Facility (STLF) (Akizuki and Ohno 2016). In this experiment, skin-color paper samples were used, referred to as "paper color chart", which are Munsell color samples selected to match the colors of real skin for different circulatory dysfunctional conditions for gender and age groups. Their spectral reflectance factor curves, however, were not close to real skins' data especially in the region longer than 580nm, and the results of visual evaluation was not satisfactory.

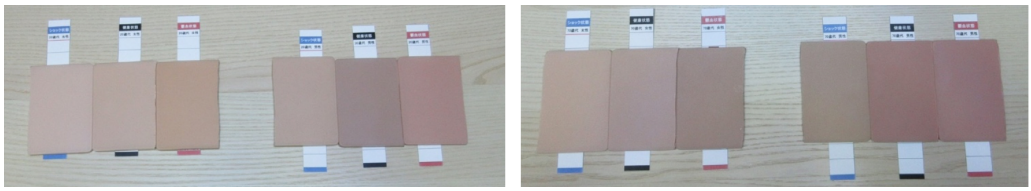
In this study, new skin model samples using urethane have been developed for different circulatory dysfunction conditions for improved spectral reflectance characteristics as well as sample's visual appearance, texture, and structures to be closer to real human skins. The characteristics of these urethane samples have been evaluated with colorimetric calculations

Akizuki, Yuki; Ohno, Yoshihiro.
"Preliminary Study on Spectral Characteristics For Identification of Skin Color Under Circulatory Dysfunction Using Artificial Skin Samples."
Paper presented at CIE 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

and by two vision experiments, to examine suitability for use in evaluating light sources for distinguishing human skin colors under circulatory dysfunction conditions. The experimental methods for evaluating light sources using these skin model samples have also been evaluated.

2 Development of Urethane Skin Model Samples and their characteristics

The developed urethane skin model samples are shown in Figure 1. They are 5 mm thick, with the size of 100 mm × 70 mm, and made of two layers of urethane providing appearance with surface texture similar to real skins, with colors matched to those of three different circulatory dysfunction conditions (healthy, shocked, and congested) for four gender/age groups (young female, young male, elderly female, and elderly male). Table 1 shows the L^*, a^*, b^* color coordinates of these samples (under D65) measured with a spectrophotometer. The expanded uncertainties (k=2) of the measured L^*, a^*, b^* values were estimated to be 0.2, 0.1, 0.1, respectively, and of ΔE^*_{ab} to be 0.2. Figure 2 shows the spectral reflectance factor data of these samples. In Table 1 and Figure 2, the data of real human skins (Akizuki et al. 2013) and data of the paper color charts (for young female's skin color) used in our previous study are also included for comparison. The spectral reflectance factor curves of the skin model samples are closer to real skin data than the paper color charts at longer than 580 nm, but they still do not match sufficiently.



Young female Young male Elderly female Elderly male

Figure 1 – Urethane skin model samples. Each set is placed in the order of shocked (left), healthy (center), congested (right).

Table 1 – Color data of real human skin, urethane skin model samples and paper color chart measured with a spectrophotometer under CIE standard illuminant D65. ΔE*ab was calculated between real human skin and urethane skin model or paper color chart.

		Young Female			Young Male			Elderly Female			Elderly Male		
		Healthy	Shocked	Congested	Healthy	Shocked	Congested	Healthy	Shocked	Congested	Healthy	Shocked	Congested
Real Human Skin	L^*	65.2	68.3	61.9	60.5	62.6	57.8	60.8	63.8	59.0	55.8	59.6	54.5
	a^*	7.4	4.1	11.1	8.8	6.5	12.6	7.9	5.0	11.5	10.5	5.8	12.3
	b^*	17.8	19.2	19.1	17.8	21.3	19.9	16.9	20.1	18.4	18.9	21.4	18.2
Urethane Skin Model Sample	L^*	64.9	69.5	61.3	59.8	62.0	58.3	61.0	61.5	58.1	54.9	59.0	54.6
	a^*	8.9	10.0	10.2	10.9	8.2	11.8	8.3	8.1	8.9	14.2	8.0	13.6
	b^*	17.0	16.6	20.8	14.8	16.5	14.4	15.7	15.9	17.2	16.1	18.4	14.1
ΔE^*_{ab}		1.7	6.6	2.0	3.8	5.1	5.6	1.3	5.8	2.9	4.8	3.7	4.4
Paper Color Chart	L^*	60.2	70.7	60.2									
	a^*	7.5	4.5	11.9									
	b^*	15.8	18.6	18.2									
ΔE^*_{ab}		5.3	2.5	2.1									

These urethane skin model samples were used as evaluation targets in the experiments by using NIST STLF. The same set of light spectra on the NIST STLF used in the previous study (Akizuki and Ohno 2016) was used. These SPDs are shown in Figure 3. The three-band spectra (labeled 6500 K LED and 2700 K LED in the figure) are called "optimized LED lights" in this paper, and they were tuned as close as possible to the theoretical SPDs that were determined to produce the largest color differences in the real skin colors in the different circulatory dysfunction conditions, within the range of D_{uv} (distance from Planckian locus on CIE u', 2/3 v' coordinates, with + sign above, - sign below Planckian locus) for acceptable white light and color rendering for lighting (D_{uv} from -0.02 to 0.01). The broadband spectra (labeled 6500 K Daylight and 2700 K Incandescent in the figure) are called "reference lights" in this paper, and they were tuned close to Daylight illuminant D65 and incandescent light at 2700 K,

respectively, at varied D_{uv} levels. Note that the peak around 550 nm of the optimized LED lights is broad (though it should ideally be narrowband) because there are no narrowband high-power LEDs available around this wavelength and this peak is one of the green phosphor LEDs used in STLF.

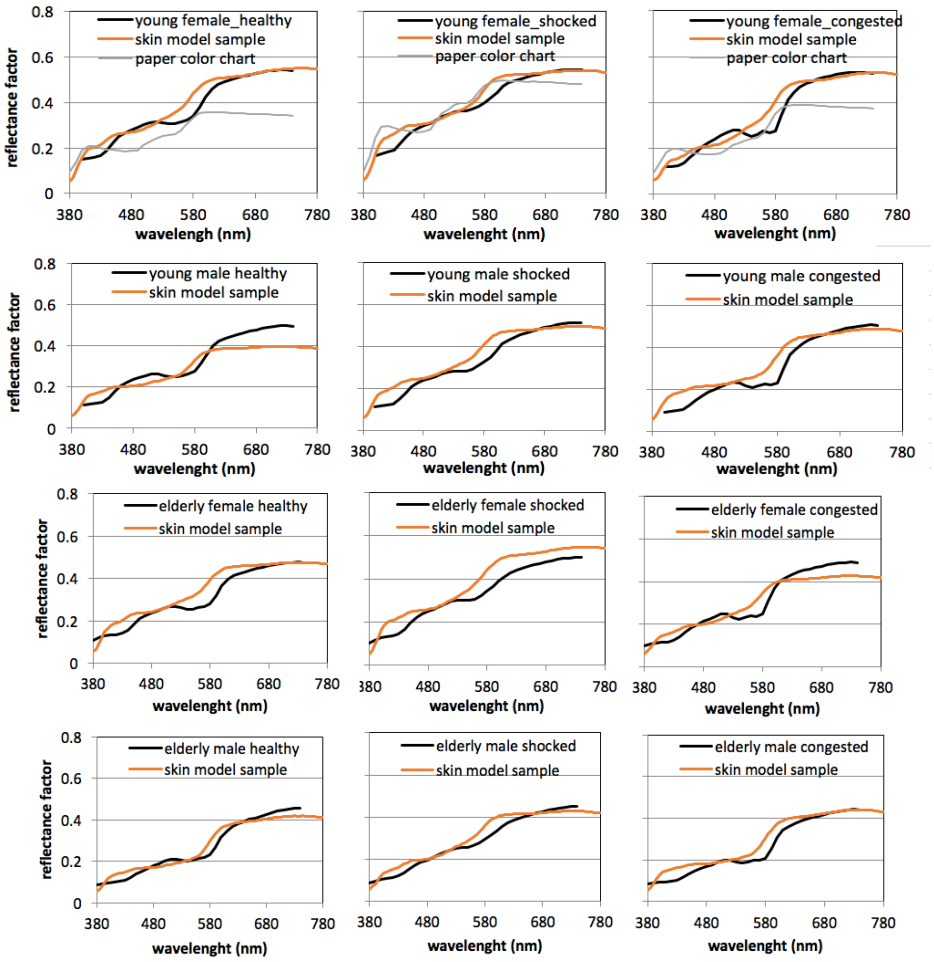


Figure 2 – Spectral reflectance factors of real human skin (Japanese), urethane skin model samples and paper color chart, for different gender and age groups.

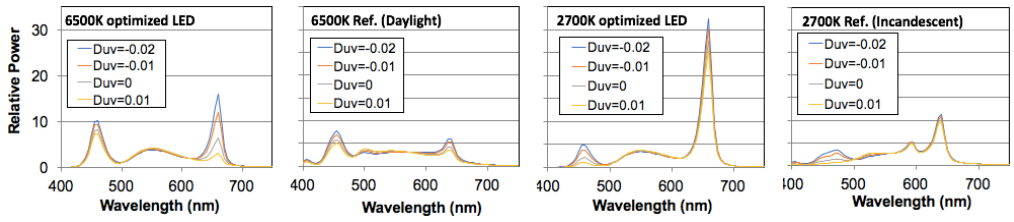


Figure 3 – Spectral power distributions of light sources used in the analysis

Figure 4 shows an example (for young female) of calculated color differences ΔE^*_{ab} between healthy and shocked conditions (written as "H-S" in the figure) or healthy and congested conditions ("H-C" in the figure), for real human skin, paper color charts and urethane skin model samples, under different SPD lighting conditions in Figure 3. Ideally, the curves for skin model

samples should be close to those for real skin. This example shows that the color difference is very large for paper chart H-S condition for all sources. The color differences for urethane sample H-S condition also have some shift from real skin, but shift is much smaller than the paper color chart. Other conditions (young male, elderly female, elderly male) showed similar range of variations of the curves. From these, we judge that urethane model samples are improved from the paper color chart (closer to real skin) in terms of relative color differences.

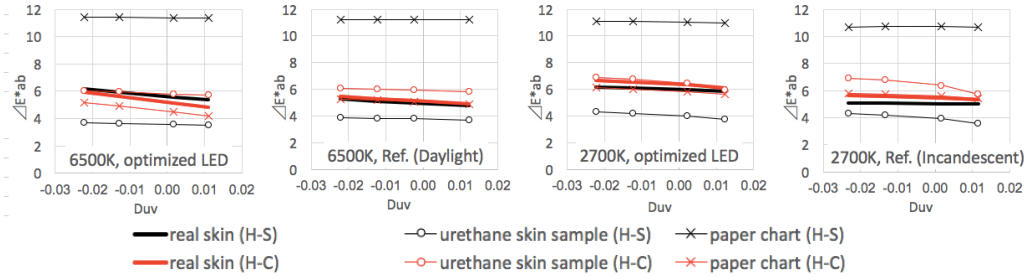


Figure 4 – Comparison of calculated color differences between healthy and shocked or congested conditions for young female's visual targets under light spectra in Figure 3.

Figure 5 (a) and (b) show calculated color differences between healthy and shocked or congested conditions, for each visual target (human real skin, urethane samples, paper samples) for each age and gender group, under the different types of SPD in Figure 3. "t-LED" and asterisk plots in Figure 5 indicates the theoretical LED by 3 narrow-band LED spectra (including 550 nm narrow peak) described in our previous paper (Akizuki and Ohno, 2016). For real human skin, as expected, the color differences under theoretical LEDs are largest in all cases. The optimized LEDs were designed using real available LEDs as used in NIST STLF to be close to the theoretical LEDs and to produce maximum color differences, therefore their results of color differences should be larger than the ones under the reference lights. This is clearly shown for real skins, but the differences for urethane samples are much smaller. Also, the optimized LEDs, in many cases, show color differences increasing when D_{uv} shifts to negative direction (toward -0.03). Some data of the urethane skin model samples and paper color charts show similar tendency, but other results show much smaller slopes of the curve. These results show that these developed urethane skin model samples need further improvements.

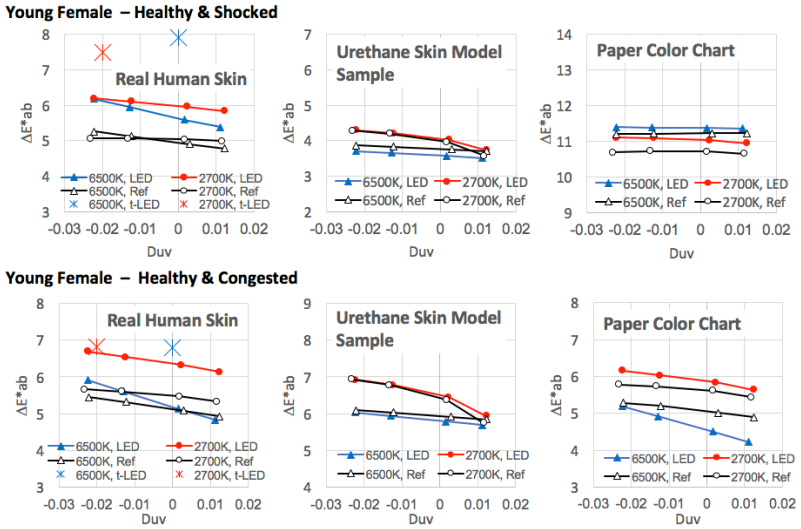


Figure 5 (a) – Calculated color differences between healthy and shocked or congested conditions, for young female, under different lighting spectra with varied D_{uv} in Figure 3 and the theoretical LED.

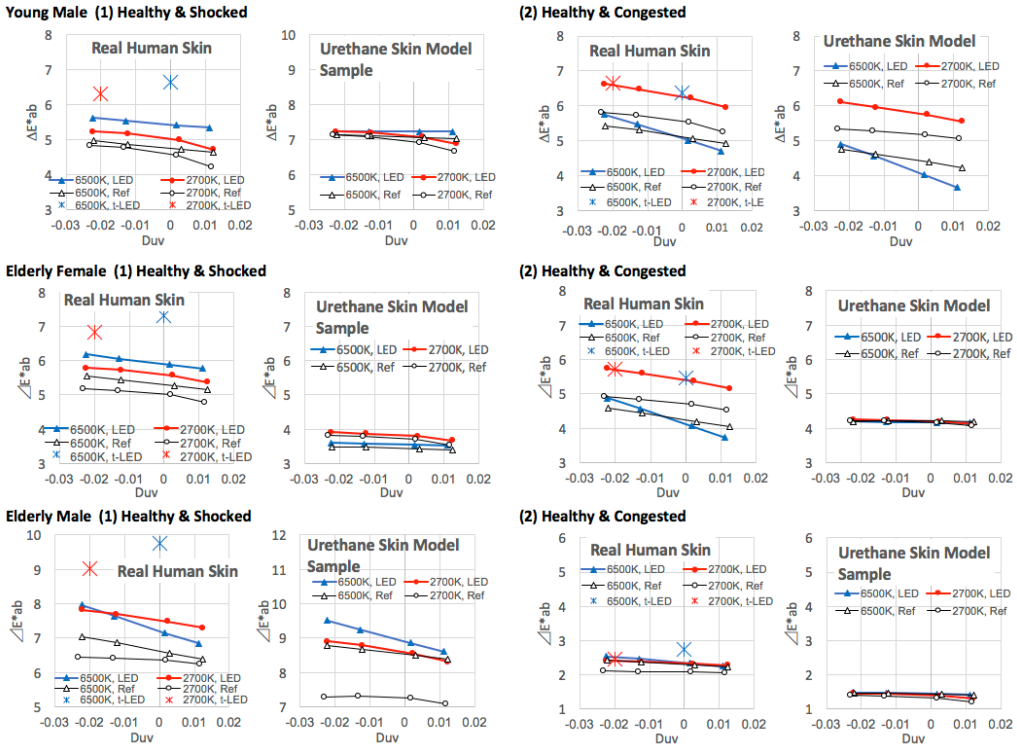


Figure 5 (b) – Calculated color differences between healthy and shocked or congested conditions, for young male and elderly female and male, under the lighting spectra with varied D_{uv} in Figure 3 and the theoretical LED.

3 Consideration of Experimental Methods for Evaluating Effective Light Sources

Although the developed urethane skin model samples need further improvements as mentioned in Section 2, these are better skin samples to simulate the circulatory dysfunction conditions than paper color charts and any other samples at this time. Therefore, we used these urethane skin model samples for consideration of the appropriate experimental methods for evaluating effective lighting spectra to distinguish the color differences of human skin under circulatory dysfunction conditions.

Two different experiments have been conducted for this purpose. Experiment-1 used the same procedures in the previous study (Akizuki and Ohno 2016), and its purpose was to judge the naturalness of lights at different D_{uv} levels, and the most effective levels of D_{uv} of the optimized LED lights for distinguishing skin model sample colors at different conditions. The optimized LED lights (three-band SPD) and reference lights (broadband SPD) on STLF shown in Figure 3 were used in the experiment. Experiment-2 directly compared naturalness and distinction of sample colors between the optimized LED lights and the reference lights and its purpose was to verify whether the optimized LED lights can distinguish different skin model sample colors better than the reference lights, at varied D_{uv} levels. In both experiments, the same D_{uv} values (from -0.02 to +0.01) and the same correlated color temperature (CCT) (2700 K and 6500 K) of the lights shown in Figure 3 were used.

3.1 Subject

The subject was one female with normal color vision in her 40s (first author). This subject was well acquainted with Experiment-1 in the previous study, and trained thoroughly. At both experiments, the subject sat on a couch placed at a room of STLF, which had a simulated interior room environment with a table, a bookshelf, and other objects. Pictures of the

experimental scene are shown in Figure 6. For facilitating the experimental operation easier by herself, the controller of STLF, the evaluation sheet, and visual targets were placed on the same table. In the experiments, the urethane skin model samples and the paper color charts were arranged side by side with three different circulatory dysfunction conditions (left-shocked, center-healthy, and right-congested) by each age and gender groups (see also Figure 1).

Figure 6 – Experimental scene

3.2 Procedure of Experiment-1

The procedure of Experimental-1 was the same as those used in the previous study (Akizuki and Ohno 2016), and the purpose was to examine naturalness of different D_{uv} lights for skin color as well as distinction of model skin samples under different D_{uv} levels.

The illuminance level on the center of the table in the STLF room was set to 200 lx and kept within ± 2.5 % during the course of experiment. At each combination of CCT and SPD, the subject was first adapted to the illumination at one end of D_{uv} (D_{uv}= 0.01 or -0.02) for five minutes in order to be adapted to the illumination completely. And then, a pair of lights, which had a D_{uv} level higher or lower than 0.005 from the adapted light (e.g., D_{uv}=0.015 and 0.005 for the adapted light at D_{uv} =0.01) was presented. The pair of lights was flipped at about three seconds interval and repeated several times, while the subject selected which light in the pair for the four questions listed below. The answers were forced choice, and if the judgement was difficult, she reported that also.

Q1: Which light presents the color of the back of her left hand more natural?

Q2: Which light presents the urethane skin model samples more natural?”

Q3: Which light presents the color differences of two combinations of the paper color charts (healthy and shocked, healthy and congested) larger?

Q4: Which light presents the color differences of two combinations of the urethane skin model samples (healthy and shocked, healthy and congested, for each group) larger?

Then, next D_{uv} was presented (e.g. D_{uv}=0) and the subject was adapted to the illumination for one minute, and the same trial with a pair of lights (D_{uv}=0.005 and -0.005 for adapted light at D_{uv}=0) was made. This was repeated for four adapted D_{uv} levels (0.01, 0, -0.01 and -0.02), which completed one run for each SPD type (6500 K optimized LED, 6500 K reference as daylight, 2700 K optimized LED, and 2700 K reference as incandescent). Then another run for the same condition was conducted in reverse order of D_{uv} (start from D_{uv}=-0,02 and ends at D_{uv}=0,01). These two runs (ascending order and descending order) were repeated for all SPD types.

3.3 Procedure of Experiment-2

The experiment-1 does not compare the distinction of colors of samples between the optimized LED lights and the reference lights. In Experiment-2, the subject compared the optimized LED lights directly with the reference lights for natural appearance of real skin and for distinction of colors of the model samples for different circulatory dysfunction conditions at different D_{uv} levels.

The purpose of Experiment-2 was to find the most effective D_{uv} conditions of the optimized LED lights than the reference lights,

The illuminance level on the table in the STLF room was set to 200 lx, too. After adaptation to the first reference light for five minutes, the optimized LED light and the reference light having the same D_{uv} and CCT were presented sequentially, and subject answered the four questions described in 3.2. If the judgement was difficult, she reported that also. Further, she answered the degree of naturalness or degree of distinction of the optimized LED light compared with reference light in five evaluation scales as shown below.

-2: definitely Light A, -1: rather Light A, 0: uncertain, 1: rather Light B, 2: definitely Light B

The D_{uv} value of the optimized LED and reference light were set to eight levels (-0.025, -0.02, -0.015, -0.01, -0.005, 0, 0.005, 0.01), and the subject was adapted to five minutes for the first light (D_{uv}= -0.025 or 0.01) and to each point afterwards for one minute before making judgment. The order of presentations of these D_{uv} levels were done in ascending and descending directions for each session as in Experiment-1.

3.4 Results of Experiment-1

Tables 2 and 3 show the raw results of Experment-1. In these tables, “ascending” means that the presented order of the D_{uv} condition was from -0.02 to 0.01 (upward in the table), and “descending” means from 0.01 to -0.02 (downward in the table). “Lower D_{uv}” means that the subject chose the light with lower D_{uv} value (chromaticity shift in pinkish direction from the adapted D_{uv} point) in the pair for the answer, and “higher D_{uv}” means that the subject chose the light with higher D_{uv} value (in yellowish direction). The asterisk mark (*) means that the subject had difficulty in judgement. Although the subject should make a binary choice “lower D_{uv}” or “higher D_{uv}” under difficult situation mandatorily, we considered the difficulty answer meant “uncertain” and did not treat the D_{uv} direction.

Table 2 shows the raw results of natural appearance (see 3.2 Q1 and Q2). The selected D_{uv} direction (lower or higher D_{uv}) tend to change from lower to higher in ascending direction, and from higher to lower D_{uv} in descending direction. The D_{uv} level where this cross-over occurs is considered to be most natural D_{uv} level (Ohno and Oh, 2016), and it is around D_{uv}=-0.005 for real skin color, and it is around D_{uv}=0 for urethane skin model samples, which is close to the results for real skin.

Table 2 – Row results of Experiment-1 on natural appearance.

(1) Natural-appearance of hand's skin color

	6500K LED		6500K Ref. (Daylight)		2700K LED		2700K Ref. (Incandescent)	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0.01	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
0	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	**Higher Duv**	lower Duv	lower Duv
-0.01	*	lower Duv	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**
-0.02	**Higher Duv**	lower Duv	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**

(2) Natural-appearance of urethan skin model samples

	6500K LED		6500K Ref. (Daylight)		2700K LED		2700K Ref. (Incandescent)	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0.01	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv	*
0	lower Duv	**Higher Duv**	**Higher Duv**	**Higher Duv**	lower Duv	*	*	*
-0.01	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	*	**Higher Duv**	**Higher Duv**	**Higher Duv**
-0.02	**Higher Duv**	**Higher Duv**	**Higher Duv**	**Higher Duv**	lower Duv	**Higher Duv**	**Higher Duv**	**Higher Duv**

Table 3 shows an example of raw results of distinction of urethane skin model samples for young female in different circulatory dysfunction conditions (see 3.2 Q3 and Q4). In Figure 5 presented in section 2, some graphs show larger falling slopes of the curves of color difference vs. D_{uv}, while other graphs show fairly flat curves. If the slope is large, lights with lower D_{uv} produce larger color differences and “lower D_{uv}” should be chosen in this experiment. If the curve is flat, the judgement is difficult and “higher Duv” and “lower D_{uv}” answers should be mixed, and “difficult” answer should increase.

Table 3 – Example raw results of Experiment-1: distinction of colors of urethane skin model sample for young female

	6500K LED		6500K Ref. (Daylight)		2700K LED		2700K Ref. (Incandescent)	
Duv	ascending	descending	ascending	descending	ascending	descending	ascending	descending
0.01	lower Duv	lower Duv	*	lower Duv	lower Duv	lower Duv	lower Duv	lower Duv
0	*	**Higher Duv**	lower Duv	**Higher Duv**	lower Duv	lower Duv	*	lower Duv
−0.01	lower Duv	lower Duv	**Higher Duv**	**Higher Duv**	lower Duv	*	lower Duv	lower Duv
−0.02	lower Duv	*	**Higher Duv**	**Higher Duv**	lower Duv	lower Duv	lower Duv	lower Duv

To check this consistency of subject evaluation, the frequency distribution of subject answers was calculated. Figure 7 shows the ratio of "lower D_{uv}" (black) and "uncertain (*)" (white) to the total number of answers (eight) for each condition. These results show that the subject evaluated the changes of color differences with D_{uv} correctly or not. If the slope of the curve in Figure 5 is large, the "lower Duv" ration should be high. If the slope of the curve in Figure 5 is flat, the "lower Duv" ratio should be low (less than 50 %) and ratio for uncertain" should be high. For example, many of the curves for 6500 K reference for urethane skin model samples in Figure 5 are nearly flat, which is consistent with the 6500 K reference results (white bar dominant) in Figure 7. Also, many of 2700 K LED in Figure 7 for urethane skin model samples show larger slopes of the curves, which is consistent with the 2700 K LED results (high black bars) in Figure 7. However, these results do not provide optimum Duv levels for distinction of color differences. There are also some cases where results are not consistent between calculation (Figure 5) and visual evaluation (Figure 7). Experiment-1 is considered to be useful to find suitable Duv levels for natural appearance of skin, but not for evaluating distinction of color differences of model samples.

3.5 Results of Experiment-2

Table 4 shows an example of raw results of Experiment-2. The upper part of each table (reference or LED) means that subject chose the reference light or the optimized LED as the answer. Asterisk mark (*) means that judgement was difficult. In the lower part of each table, the numbers (1, -1, etc.) show the evaluation value in the scale for degree of naturalness or distinction of the optimized LED light compared with the reference light in five evaluation scales (see 3.3). These results are consistent with the results in the upper part of the table (simple choice of which light) and showing more detailed information.

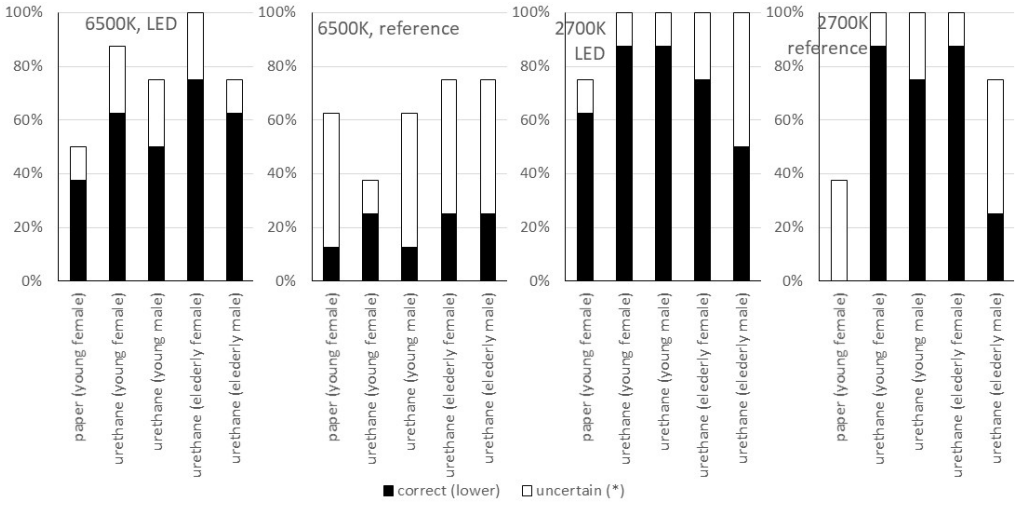


Figure 7 – Summary results about distinction evaluations of Experiment-1

Table 4 – Example raw results of Experiment-2 (naturalness of hand' skin color)

Duv		−0.025	−0.02	−0.015	−0.01	−0.005	0	0.005	0.01
6500K	ascending	reference	LED	reference	*	*	reference	reference	reference
	descending	reference	reference	reference	*	reference	reference	reference	reference
2700K	ascending	reference	reference	reference	reference	reference	reference	reference	reference
	descending	reference	reference	reference	reference	reference	reference	reference	reference

Duv		−0.025	−0.02	−0.015	−0.01	−0.005	0	0.005	0.01
6500K	ascending	−1	1	−1	0	0	−1	−2	−2
	descending	−2	−1	−1	0	−1	−1	−2	−2
2700K	ascending	−1	−2	−1	−2	−2	−2	−2	−2
	descending	−2	−2	−2	−2	−2	−2	−2	−1

Judging from the comparison between ascending order and descending order, it seems that Experiment-2 (the direct comparison of the optimized LED to reference) produced more stable results than Experiment-1 (the judgment of color differences of the samples between different D_{uv} levels).

Figure 8 (1) shows the plots of the degree of naturalness and Figure 8 (2) to (6) show distinction of the sample colors in the five number scales (as the example shown in the lower part of Table 4) for comparing the optimized LED light and the reference light at each D_{uv} level. The results of ascending and descending results were averaged in these plots. Q2 (naturalness of the model samples) was not included in Experiment-2 because Experiment-1 already showed that the results of urethane samples were close to the results of real skin, and it was also considered that judgement of naturalness of a piece of artificial sample in a scale would be difficult. In the case of natural appearance of real hand's skin color as shown in Figure 8 (1), the reference lights produced more natural appearance than the optimized LED lights at any D_{uv} levels. This may be due to the fact that the optimized LED lights have high chroma increase (in red-green direction) and skin color appears slightly more reddish than the reference light. This is considered as a trade-off point with increased color contrast with such chroma-enhanced sources.

In the case of distinction of paper color charts shown in Figure 8 (2), the optimized LED lights (both 6500 K and 2700 K) produced higher distinction than the reference. These results are mostly consistent with the calculation data in Figure 5, except 6500 K LED for healthy and congested, predicting opposite result. In this case, the healthy and shocked comparison shows consistent theoretical data. The results could have changed depending on which combination of samples the subject paid attention to when making judgment.

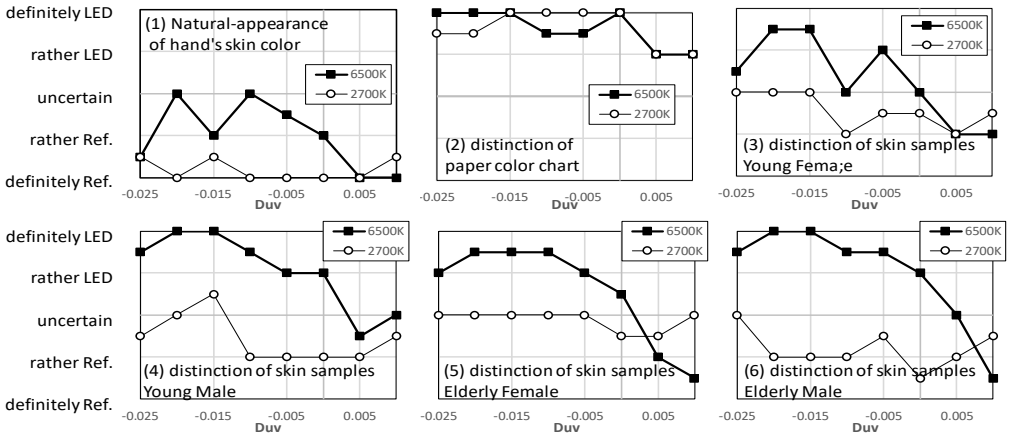


Figure 8– Summary results of Experiment-2

In the case of the distinction of urethane skin model samples shown in Figure 8 (3) to (6), the two CCT conditions had different tendency. For 6500 K, LED lights were chosen for higher distinction of samples, especially at negative Duv levels. For 2700 K, the results were less certain and reference lights were rather chosen for higher distinction. These results do not agree well with the results shown in Figure 5, where differences (between LED lights and reference lights) are very small in many cases, and 2700 K

LEDs rather show higher color differences (young male – healthy and congested, also elderly male – healthy and shocked) than 6500 K LED lights.

It should be noted that the results in Figure 5 show very different tendencies between the two graphs for each gender/age group, healthy and shocked, and healthy and congested, particularly in young male and elderly male. Depending on which two samples (shocked and healthy, or congested and healthy) the subject pay attention to, the results could be different. This is considered to be one of the reasons for inconsistent results. Also, judging from the data presented in Figure 5, the calculated color differences (between LED lights and reference lights) are very small in many cases and judgment must be very difficult.

4 Conclusion

Urethane skin model samples were developed to be possibly used for visual evaluation of light sources for distinction of circulatory dysfunction conditions of skin color. These are better skin model samples than the paper color charts. The samples are found useful for evaluation of naturalness of light sources. The spectral reflectance characteristics are found not sufficiently matched to real human skin for these conditions. However, the visual evaluation of distinction of color differences did not agree well with the calculation results. Therefore, skin model samples need to be improved for spectral reflectance curves matched closer to those of real human skin.

It is demonstrated that Experiment-2 provided more information on evaluation of the LED sources including comparison to the reference sources (it is not covered in Experiment-1), and found to be a useful evaluation method.

There were large variations of results in this experiment, so more refined experimental methods are needed, e.g., improved method of presenting visual target separately for shocked and healthy, and for congested and healthy conditions. In both experiments in this study, pair of lights (optimized LED and reference) were presented alternately in time sequence. Another experimental method to compare two lights simultaneously by haploscopic vision may further improve reproducibility of results.

Experiments were conducted only with one subject to evaluate experimental methods. Further experiments with a larger number of subjects are needed for evaluation of LED light sources for use in practical rescue sites.

For a longer term, it is also desired to have narrow-band green emission to make the theoretical LED to realize more effective light sources for distinguishing different circulatory dysfunction conditions of skin color.

Acknowledgments

This research was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP16K06607 (Program Leader: Yuki Akizuki). The authors would like to thank Mr. Takashi Suguro of Murakami Color Research Laboratory and Mr. Yasuhide Takane and Mr. Nobuo Kihara of Beaulax Co. for their cooperation in the development of urethane skin model samples.

References

AKIZUKI, Y. TOMOHIRO, J. WAKASUGI, M. YOSHIMURA, A. TAKASHINA, K. 2013. Creation of Colour chart database of disaster victim's skin - Measurement of quasi-skins in shock and congested state by healthy young subjects, *Proc. 12th International Colour Association*, 4pages.

AKIZUKI, Y., OHNO, Y. 2016. Study on Spectral Characteristics for Identification of Skin Color of Injured Japanese Persons at Disaster Sites, CIE x043:2016, pp.179-189.

DAVIS, W., OHNO, Y. 2010. Color Quality Scale. Optical Engineering, 49 (3) 033602, pp.1-16.

OHNO, Y., OH, S. 2016. Vision Experiment II on White Light Chromaticity for Lighting, CIE x042:2016, pp.175-184.

OHNO, Y., FEIN, M. 2014. Vision Experiment on Acceptable and Preferred White Light Chromaticity for Lighting, CIE x039:2014, pp.192-199.

GLOBAL INTERLABORATORY COMPARISON OF GONIOPHOTOMETER MEASUREMENTS USING LIGHT EMITTING DIODE ARTEFACTS

Jeon, S.K.[1], Ohno, Y.[2], Nara, K.[3], Scholand, M. J.[4], Dubard, J.[5], Borg, N.[4], Boughey, D.[6], Bennich, P.[7]

[1]Korea Institute of Lighting Technology, Republic of Korea, [2]National Institute of Standards and Technology, USA, [3] IA Japan/NITE, Japan, [4]IEA 4E SSL Annex, [5]Laboratoire Nationale de métrologie et d'Essais [6]Department of the Environment and Energy, Australia, [7]Swedish Energy Agency, Sweden

ssl.annex@gmail.com

Abstract

The International Energy Agency's (IEA) Energy Efficient End-use Equipment (4E) Solid-State Lighting (SSL) Annex is conducting a global interlaboratory comparison for goniophotometers for measuring light-emitting diode (LED) lighting products, called Interlaboratory Comparison 2017 (IC 2017). It is not only a technical study for comparison of various goniophotometric quantities of LED luminaires and lamps, it is designed in compliance with ISO/IEC 17043 to serve as proficiency testing for those laboratories seeking for testing laboratory accreditation for goniophotometers using CIE S 025 (or EN 13032-4) or other regional test methods. Participating laboratories will measure a set of four LED product artefacts: (1) a narrow-beam LED directional lamp, (2) an indoor LED panel luminaire, (3) an indoor linear LED batten luminaire, and (4) an LED street lighting luminaire. This paper describes the technical background of the design of IC 2017 and the Technical Protocol including the formula for data analysis.

Keywords: Interlaboratory Comparison; Goniophotometer; LED Luminaire; Proficiency Test; CIE S 025; IEA 4E SSL Annex; LED Metrology; Accreditation; IC 2017; Laboratory Accreditation

1 Introduction

Policy makers around the world need repeatable and reliable performance testing of solid-state lighting products. Testing is the foundation of any market transformation programme intended to move toward high quality, more environmentally responsible products. Robust measurement of the performance of LED lamps and luminaires has been challenging, and many laboratories have experienced difficulties in calibrating their equipment and establishing procedures that generate reliable and consistent test results.

Building on the experience of the interlaboratory comparison 2013 (IC 2013) where 110 labs from around the world were compared for measurements of solid state lighting (SSL) products (SSL Annex 2014), the IEA 4E SSL Annex launched Interlaboratory Comparison (IC 2017) for the measurement of SSL products by goniophotometers (SSL Annex 2017a). IC 2017 is organised to compare measurements of LED luminaires as well as a narrow-beam LED lamp, neither of which were covered in IC 2013. IC 2017 will study the equivalence of measurements by different types of goniophotometers, not only the traditional far-field types but also near-field goniophotometers, and investigate the measurement variations of participating laboratories using goniophotometers to measure SSL products.

The SSL Annex is supporting IC 2017 as part of its intergovernmental work promoting harmonised approaches to LED performance specifications and testing. The IC will use CIE S 025 (Test Method for LED lamps, LED luminaires, and LED modules) (CIE 2015) as the test method, and is intended to build capacity in test laboratories in support of the use of CIE S 025.

Ohno, Yoshihiro.
"Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

IC 2017 is also designed to be compliant with ISO/IEC 17043 (ISO/IEC 2010) to serve as a proficiency test for SSL testing accreditation programmes that recognise this comparison, as was done in IC 2013.

2 Structure of IC 2017

All types of goniophotometers covered in CIE S 025 are eligible to participate in IC 2017, including goniophotometers with photometer head, gonio-spectroradiometers, far-field and near-field goniophotometers. Near-field goniophotometers and non-standard goniophotometers that rotate the operating position of the luminaire with a correction technique (allowed in CIE S 025) will also be covered in this interlaboratory comparison, and their results may be used for the validation requirement in CIE S 025 for such goniophotometers or to assist in an accreditation application. Sphere-spectroradiometers for colour measurements, used in combination with a goniophotometer with photometer head, are also included.

IC 2017 is being conducted as a star-type comparison between one of many participating labs and one of two Nucleus Labs which provide the assigned values. An artefact set consisting of the three luminaires and one narrow-beam lamp is measured by the Nucleus Lab and then sent to a participant for measurement. When the participant has completed and submitted their data, they return the artefact set to the same Nucleus Lab for a second measurement. The assigned values are determined from the Nucleus Lab measurements before and after the measurements by participant, with the correction factors described in section 3.1 of this paper.

The measurement rounds for IC 2017 will start in October 2017 and will be completed by May 2018. There will be three sequential measurement rounds, to which participants will be assigned after the fee is paid, in the order of payment received, or they can request a later round if they wish. The table below provides an overview of the timeline of IC 2017.

Table 1 – Timeline of Interlaboratory Comparison 2017 (IC 2017)

Item	Date
Announcement and opening of application period for participants	30 Jun 2017
Closure of the application period	30 Sep 2017
Round 1 Measurements conducted by the participants	Oct 2017 - Nov 2017
Round 2 Measurements conducted by the participants	Jan 2017 - Feb 2018
Round 3 Measurements conducted by the participants	Apr 2018 – May 2018
Individual Test Reports (ITR) issued to participants	3 months after each testing round
Draft Final Report of IC 2017 issued to participants	November 2018
Final Report of IC 2017 issued to the public	March 2019

2.1 Test Method – CIE S 025

IC 2017 is designed to be compliant with ISO/IEC 17043 to serve as a proficiency test for SSL testing accreditation programmes that recognise this comparison. IC 2017 will use CIE S 025 (Test Method for LED lamps, LED luminaires, and LED modules) as the test method.

If recognised by accreditation bodies, the participant's individual test report may be used as a proficiency test (PT) for CIE S 025. The technical requirements in CIE S 025 encompasses many regional and national test methods such as EN 13032-4 (EN 2015) (European standard, equivalent to CIE S 025); LM-79 (North America); KS C 7653 and KS C 7651 (Korea), and JIS C7801 and JIS C8105-5 (Japan). If recognised by accreditation bodies, the participant's test

Ohno, Yoshihiro.
"Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

report could also be used as the objective evidence of the competence of the laboratory for these test methods and possibly also those in China and elsewhere. Please see the IEA 4E SSL Annex's report "Application Study of CIE S 025/E:2015" (SSL Annex 2017b) which provides more information on equivalencies and provides a table in the Annex which compares the differences between the various test standards.

2.2 Artefact Selection and Quantities Measured

The full set of comparison artefacts consists of four LED lighting products as shown in Figure 1. One set of these four artefacts will be sent to each participant in two rigid shipping containers.

ART-1 ART-2 ART-3 ART-4

Figure –1 The four artefact types used in IC 2017

The artefacts were selected as typical indoor and outdoor luminaires (and a lamp) in the current market as below.

- ART-1 is a typical MR-16 narrow beam lamp especially to compare measurements of beam angle, centre beam intensity, and partial flux;
- ART-2 is a typical indoor planar luminaire with broad (near-Lambertian) distribution;
- ART-3 is another indoor luminaire with small upward light emission especially to test capability for measurement of upward luminous flux; and
- ART-4 is a typical street lighting luminaire having significantly asymmetric intensity distributions in the horizontal plane, where a product model with fairly low power factor was chosen.

The rated power of each artefact is 7.5 W, 40 W, 20 W, 20 W, from ART-1, 2, 3, 4, respectively. All of the artefacts will be seasoned for 200 h and tested for reproducibility (three measurements, each cold start), and calibrated by the Nucleus Labs following CIE S 025, before shipping to participants.

Several sets of these artefacts have been tested at KILT for reproducibility in different dates. Luminous flux of these artefacts reproduced mostly to within ±0.3 %, and at the worst case ±1 %, chromaticity *u*', *v*', within ±0.0003 and at the worst case ±0.0006.

A representative set of these artefacts were tested for temperature sensitivity, input voltage sensitivity, and current waveform, to provide information for uncertainty budget. The temperature coefficient for luminous flux ranged from -0.2 % to -0.8 % per °C, that for chromaticity u', v' was within 0.0002 per °C. The input voltage coefficient for luminous flux was within 0.2 % per 1% change of voltage, and that for chromaticity u', v' was within 0.0002 per 1 % change of voltage. The power factors of these products ranged from around 0.7 to 0.97 and current wave forms of all products did not show spiky peaks and high frequency components but one of the artefacts showed significant phase shift.

2.3 Measurement Quantities

Table 2 and Table 3 show the lists of the measurement quantities to be measured and compared in this IC test. The quantities in Table 2 are the same as those measured in IC 2013. The reported results for quantities in Table 2 will be analysed to assess the laboratory performance. Therefore, the report submitted to the participant as Individual Test Report may

Ohno, Yoshihiro.
"Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

be readily used as objective evidence in applying for accreditation. The quantities in Table 3 are goniophotometric quantities. The results for quantities in Table 3 will be used mainly for the purposes of a technical study. While we request all participants to measure and report results of all the quantities in Table 3, if any participants have difficulty, it would be acceptable not to report one or more quantities in Table 3.

Table 2 – General Quantities (for all Artefacts)

No.	Quantity	Units
1	Total luminous flux	lm
2	Luminous efficacy	lm/W
3	Chromaticity coordinate (u', v') – spatially averaged	1
4	Correlated colour temperature (CCT) – spatially averaged	K
5	Colour rendering index (CRI) Ra – spatially averaged	1
6	Active power	W
7	Root-mean-square (RMS) current	A
8	Power factor	1

Table 3 – Goniophotometric Quantities – Measure those Marked with an X

No.	Quantity	Artefact Identifier			
		ART-1	ART-2	ART-3	ART-4
9	Luminous intensity distribution (cd)	**X**	**X**	**X**	**X**
	C angle (horizontal)	2 planes only (0°/180°, 90°/270°)**			
	Vertical angle range and interval	0-90°, 1° step	0-90°, 5° step	0-180°, 5° step	0-90°, 1° step *and* 90°-180°, 10° step
10	Partial luminous flux (lm) for cone angle 15°	**X**			
11	Street light partial flux (lm)				
	Forward light				**X**
	Back light				**X**
	Up-light				**X**
12	Beam angle (avg. of 2 planes)	**X**			
13	Central beam intensity	**X**			
14	Angular spatial colour uniformity	**X**		**X**	

** These C angle planes are only for reporting luminous intensity distributions. To measure total luminous flux and partial luminous flux, goniophotometer must be scanned with C angle intervals much smaller than 90°. The vertical angle steps may also have to be smaller.

2.4 Nucleus Laboratories

IC 2017 is led by the National Institute of Standards and Testing (NIST, USA) with two Nucleus Laboratories – the Korea Institute of Lighting Technology (KILT, Korea) and the Laboratoire National de métrologie et d'Essais (LNE, France) serving as Nucleus Laboratories (i.e., pilot laboratories) for the comparison. Two Nucleus Laboratories are used in this IC to share the workload and reduce cost and time for shipping the artefacts to laboratories around the world. These Nucleus labs have their own scale of optical radiation quantities traceable to the SI, and also established measurement equivalence with NIST for measurement of total spectral radiant flux. Comparison of measurements of the LED artefacts by these Nucleus Labs is in progress (see 3.1).

KILT is the Korean Industrial Standards (KS) certification body and has been accredited for ISO/IEC 17025 by Korea Laboratory Accreditation Scheme (KOLAS) for optical radiation measurements. KILT has been also accredited as the KS testing laboratory from Korean Agency for Technology and Standards (KATS) and high efficiency certification programme testing lab from Korea Energy Agency (KEA) for photometry, and has developed the safety and photometric performance requirement for KS as a Cooperation Organisation for Standards Development (COSD). KILT also provides testing services for Energy Star programme as an EPA-recognised lab.

LNE is the National Metrology Institute for France, maintaining photometric and radiometric units (such as lumen, candela, watt) and disseminate standards for luminous flux, luminous intensity, spectral irradiance, etc. Their calibration and measurement capabilities (CMC) for many photometric and radiometric quantities are certified and published by International Committee of Weights and Measures (CIPM) under the framework of CIPM Mutual Recognition Arrangement (MRA) (CIPM 1999). LNE is accredited for ISO/IEC 17025 by COFRAC (French Accreditation Body) on a wide variety of calibration services in photometry and radiometry. LNE also provides testing services for LED lighting products compliant with CIE S 025 and is accredited for ISO/IEC 17025 by COFRAC for this activity.

Both KILT and LNE maintain a far-field goniophotometer with a photometer head + spectroradiometer (gonio-spectroradiometer) and also an integrating sphere system (2 m diameter) with a spectroradiometer (sphere-spectroradiometer). For IC 2017, both laboratories will use the goniophotometer with a photometer head for luminous flux and luminous intensity distribution, and a sphere-spectroradiometer for colour quantities (spatially averaged), and gonio-spectroradiometer for colour uniformity. The photometric distances for luminous intensity distribution are 12.5 m (KILT) and 25 m (LNE).

KILT also serves as the organising laboratory in this comparison, responsible for receiving purchased artefacts, seasoning and testing them, and shipping them to the other Nucleus laboratory. KILT also serves as the pilot laboratory for Nucleus laboratory comparison.

IC 2017 was carefully designed to be in compliance with ISO/IEC 17043, “Conformity assessment – General requirements for proficiency testing”. This was done to ensure that the work would be acceptable to accreditation bodies around the world, as evidence of proficiency in testing for CIE S 025. The two Nucleus Laboratories possess their own primary measurement standards and have developed and validated a measurement method for SSL testing. Their associated measurement services have been either accredited in the framework of International Laboratory Accreditation Cooperation Mutual Recognition Arrangement (ILAC/MRA) (case of KILT) or registered to Appendix-C of CIPM/MRA with peer review against ISO/IEC 17025 and measurement uncertainties (case of LNE). Therefore, the basic competence of Nucleus Laboratories related to the measurements in this IC was well-established.

3 Technical Protocol

A technical protocol was drafted to provide instructions to all participants for the methodology to follow in IC 2017. In this section of the paper, we present the calculation methods that are used in the Nucleus Laboratory comparison, the determination of assigned values for IC 2017,

Ohno, Yoshihiro.
”Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts.”
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

the importance of participant's uncertainty reporting and the evaluation of the performance, i.e., the E_n number (defined in ISO/IEC 17043) and the *z'* score (defined in ISO 13528).

3.1 Nucleus Laboratory Comparison

Measurements by the two Nucleus Labs are compared to establish the equivalence between them. The comparison measurements are in progress at the time this paper is written. Measurements are made for two sets of four artefacts (eight in total) for all quantities. Due to the expected measurement variations between the measurement results of the two Nucleus Labs, it was decided to use the weighted mean of their results to establish the Assigned Values for each quantity and each artefact, to which all relevant participants results will be compared. This subsection describes how the Assigned values will be calculated.

First, we calculate the weighted mean of the two Nucleus Laboratory measurement results x_1 and x_2 of each quantity based on the measurement uncertainties of each Nucleus Laboratory for the quantity:

$$\bar{x} = x_1 \cdot \left[u^{-2}(x_1) \Big/ \left(u^{-2}(x_1) + u^{-2}(x_2) \right) \right] + x_2 \cdot \left[u^{-2}(x_2) \Big/ \left(u^{-2}(x_1) + u^{-2}(x_2) \right) \right] \quad (1)$$

where:

$u(x_1)$ is the standard uncertainty of Nucleus Lab 1's measurement for quantity x;

$u(x_2)$ is the standard uncertainty of Nucleus Lab 2's measurement for quantity x;

And the uncertainty of this weighted mean is given by:

$$u(\bar{x}) = \left[u^{-2}(x_1) + u^{-2}(x_2) \right]^{-\frac{1}{2}} \quad (2)$$

Then, for quantities such as luminous flux that use relative uncertainties, the correction factors are calculated for Nucleus Labs 1 and 2 by:

$$c_1 = \frac{\bar{x}}{x_1},\ c_2 = \frac{\bar{x}}{x_2} \quad (3)$$

For quantities that use absolute uncertainties (e.g., chromaticity coordinates *u', v'*), the correction factors for Nucleus Labs 1 and 2 for the quantity are calculated by

$$d_1 = \bar{x} - x_1,\ d_2 = \bar{x} - x_2 \quad (4)$$

The Nucleus Lab comparison is using two sets of artefacts to evaluate the individual variations of the products within the same artefact type of these correction factors. The correction factors for each quantity and for each artefact type will be obtained as an average of measurements of the two artefacts. Variation of correction factors between the two artefacts will be included in the uncertainty of the correction factors. The comparison results will be reported in the Nucleus Laboratory comparison report, which will be published before the first measurement round starts in October.

3.2 Assigned Values for IC 2017

The assigned values for IC 2017 measurement rounds will be determined for each quantity and each artefact by measurements taken from one of the two Nucleus Laboratories. For quantities that use relative uncertainties (e.g., total luminous flux), the assigned values X_{c1} and X_{c2}, in the comparison run by Nucleus Lab 1 and Nucleus Lab 2, are calculated from measured results X_1 and X_2 by Nucleus Labs 1 and 2 respectively, for each quantity and each artefact using the following equation:

$$X_{c1} = X_1 \cdot c_1,\ X_{c2} = X_2 \cdot c_2 \tag{5}$$

For quantities that that use absolute uncertainties (e.g., chromaticity coordinates *u', v'*), the assigned values X_{c1} and X_{c2} are calculated by:

$$X_{c1} = X_1 + d_1,\ X_{c2} = X_2 + d_2 \tag{6}$$

Finally, the uncertainties of the assigned values from the two Nucleus Labs are calculated in a similar way, with the following calculation for Nucleus Lab 1:

$$u(X_{c1}) \approx \left[u^2(\overline{x}) + u_A^2(x_1) + u_A^2(X_1) \right]^{1/2} \tag{7}$$

where:

$u(\overline{x})$ is calculated from equation (2) above;

$u_A(x_1)$ is the type A component of the uncertainty in x_1 (Nucleus Laboratory comparison result), and

$u_A(X_1)$ is the type A component of the uncertainty in X_1 (IC 2017 measurement round result).

The uncertainty of the Assigned Values from Nucleus Lab 2 is calculated using the same methodology. This approach will ensure that the measurements by the two Nucleus Labs are equal in principle, and all participants' results are comparable to a consistent set of Assigned Values.

3.3 Uncertainty Reporting by the Participants

Participants will report their measurement results using the Results Report Form (an Excel spreadsheet) within four weeks of receiving the artefacts, except for labs who participate with more than one goniophotometer, which are given one extra week for each additional goniophotometer, up to a maximum of two additional weeks.

IC 2017 requires uncertainty statements because they are required by CIE S 025. The uncertainty of each measured quantity shall be expressed in expanded uncertainty with a confidence interval of approximately 95 % or a coverage factor *k*=2 (ISO/IEC 2008). Reporting uncertainty on the quantities listed in Table 2 is especially important if a laboratory intends to apply for accreditation. Reporting uncertainty on the quantities listed in Table 3 is requested as well, however if the uncertainty budget of those quantities have not been established by the participant, it would be acceptable to leave some uncertainty fields blank.

It is important to note that some regional and national test methods do not require uncertainty statements, thus if a participant is seeking accreditation to such a test method, then reporting uncertainty is not required for their proficiency test (PT). However, we ask all participants to report uncertainties of all measured quantities as much as possible.

3.4 Performance Evaluation

The Nucleus Laboratories provide the assigned values. The criteria used to analyse and evaluate the performance are given by the E_n number (defined in ISO/IEC 17043) and the *z'* score (defined in ISO 13528). The E_n numbers and *z'* scores will be calculated for the measurement quantities given in Table 2, but not for the quantities in Table 3 because there are not enough data to determine generic measurement uncertainties. The E_n numbers can be calculated if the uncertainty values are reported by a participating laboratory.

The *z'* score is calculated for all results, and is determined by:

Ohno, Yoshihiro.
"Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

$$z' = \frac{x - X}{\sqrt{\hat{\sigma} + u_X^2 + u_{\text{drift}}^2}} \tag{8}$$

where:

$\hat{\sigma}$ is the Standard Deviation for Proficiency Assessment (SDPA) value

In IC 2017, the SDPA value represents the generic standard uncertainty of a participant's measurement. To estimate such generic uncertainties, the Robust Standard Deviations of the goniophotometer measurements will be extracted from the IC 2013 study and used for the quantities listed in Table 2. (Note that the Robust Standard Deviation calculation is described in Algorithm A, Annex C, ISO 13528 (ISO, 2015), offering a standard deviation calculation method that minimises the effect of outlier data points). The measurement variations in IC 2017 for the quantities listed in Table 2 are expected to be similar to those found in IC 2013, even though the IC 2013 data are only based on measurement of LED lamps.

The u_x represents the standard uncertainty of the assigned value as given in equation (7). The u_{drift} is the uncertainty contribution from the expected artefact drifts (limited to be within 0.8 x SDPA) and calculated by:

$$u_{\text{drift}} = \frac{0.8 \cdot \hat{\sigma}}{2\sqrt{3}} \tag{9}$$

The values of $\hat{\sigma}$ and u_x will be included in the Nucleus Comparison Report, published prior to the first round of participant measurements. Although the final judgment whether to accept results will depend on the accreditation bodies, in general:

- If the value of $|E_n|$ > 1.0 is calculated, this is considered unsatisfactory;
- If the value 2.0 < $|z'|$ < 3.0 is calculated, this is considered questionable; and
- If the value $|z'|$ ≥ 3.0 is calculated, this is considered unsatisfactory.

The E_n numbers (if available) and the z' scores of all participants will be reported in the Individual Test Reports (ITR) and the final report in the event that they might be used by accreditation bodies.

The use of E_n numbers or z' scores or both will be determined by the accreditation bodies (it also depends on which test standard they refer to), but if an E_n number is used, participants' reported uncertainties affect the E_n numbers and thus judgment for the competence of the laboratory applying for accreditation. Currently, the uncertainties of the participant are not used in many conformity assessment programmes, however the z' score is directly based on the deviation of those results and is more commonly used in the conformity assessment of product testing.

4 Reporting

An Individual Test Report (ITR) will be prepared and sent to each of the participant laboratories within three months of their test results being received by the Nucleus Laboratory. ITRs are kept strictly confidential in IEA 4E SSL Annex, and the participant shall also keep the information on the assigned values confidential until the all rounds are finished. The ITR will present the participants' measured results on all the artefacts and quantities, and it will compare those quantities to the Assigned Values determined by the Nucleus Laboratory. The ITR will also include E_n numbers and z' scores of the quantities listed in Table 2, and the

Ohno, Yoshihiro.
"Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

comparison will include uncertainties, if available. Participants may submit the ITR as the evidence of their competence to an accreditation body or regulatory authority, in confidence.

The ITR will include the following data for all quantities and for all artefacts:

- Reported measurement results of the participant and their measurement uncertainties.
- Measurement results of the Nucleus laboratory (Assigned values) and their measurement uncertainties.
- (Relative) differences of measurement results between the participant and the Nucleus Laboratory
- *z*' scores of the comparison
- E_n numbers of the comparison

After the three measurement rounds are completed in mid 2018, a Final Report will be prepared showing all participant results and the findings of analyses on those results. Individual participants' identification will be kept confidential through the use of random lab codes, and only the participant will be informed which code in the Final Report corresponds to their lab. Participants will have the option of having their laboratory name and location included in the report. The statistical analysis of the participants' results will take place in compliance with international standard, ISO 13528. The final report will include the following data for all quantities, all artefacts, and for all participants:

- Results of Nucleus Laboratories and their uncertainties (only for one set of artefacts as an example)
- The reproducibility of artefacts before shipment and after return from participant
- Graphic presentations of all participants' results, as (relative) differences from Assigned values
- Graphic presentations of z' score of all participants
- Graphic presentations of E_n numbers of all participants
- Comparisons of results for different types of instrument (e.g., near-field vs. far-field, gonio-spectroradiometer vs. sphere-spectroradiometer for colour quantities, etc.)
- Comparisons of results for different types of artefacts
- Derivation of SDPA values and their uncertainties
- Evaluation of the consistency between reported measurement uncertainties and observed measurement variations
- Evaluation of the performance of CIE S 025 as experienced in this IC.

Conclusions

A large-scale interlaboratory comparison of goniophotometer measurements for LED lighting products (IC 2017) has been launched under IEA 4E SSL Annex, with an aim of providing technical study for variations of results in goniophotometric measurements for LED lighting products as well as providing a proficiency test for SSL laboratory accreditation programmes. The technical protocol and data analysis methods for this intercomparison were presented and discussed. This comparison is expected to provide precious data on the current state of goniophotometric measurements of various quantities for LED luminaires and lamps among numerous laboratories around the world and also to facilitate some participants applying for testing laboratory accreditations. A draft final report is expected in late 2018.

References

CIE 2015. CIE S 025/E:2015, Test Method for LED lamps, LED luminaires and LED modules

CIPM 1999. CIPM MRA, 1999. Mutual Recognition Arrangement, http://www.bipm.org/en/cipm-mra/

Ohno, Yoshihiro.
"Global Interlaboratory Comparison of Goniophotometer Measurements Using Light Emitting Diode Artefacts."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

EN 2015. EN 13032-4:2015, Measurement and presentation of photometric data of lams and luminaires, Part 4 LED lamps, modules, and Luminaires

ISO 2015. ISO 13528:2015, Statistical methods for use in proficiency testing by interlaboratory comparison

ISO/IEC 2008. ISO/IEC Guide 98-3: 2008, Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

ISO/IEC 2010. ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing

SSL Annex 2014. 2013 Interlaboratory Comparison Final Report, https://ssl.iea-4e.org/files/otherfiles/0000/0067/IC2013_Final_Report_final_10.09.2014a.pdf

SSL Annex 2017a. IEA 4E SSL Annex, Launch of IEA 4E SSL Annex 2017 Interlaboratory Comparison of Goniophotometers: http://ssl.iea-4e.org/files/otherfiles/0000/0108/IC_2017_Official_Announcement_-_Final.pdf

SSL Annex 2017b, SSL Test Methods Comparison Table; see Appendix 2 of http://ssl.iea-4e.org/files/otherfiles/0000/0110/Task_1_Application_Study_-_Final_Report.pdf

VISION EXPERIMENT ON PERCEPTION OF CORRELATED COLOUR TEMPERATURE

KWAK, Y.[1], HA, H.[1], OHNO, Y.[2]
[1] Ulsan National Institute of Science and Technology, Republic of Korea
[2] National Institute of Standards and Technology, Gaithersburg, USA
yskwak@unist.ac.kr

Abstract

The correlated colour temperature (CCT) is defined by the closest point on the Planckian locus from a light source on the CIE 1960 (*u, v*) diagram (now obsolete), while the current standard uniform colour space is the CIE 1976 (*u', v'*) diagram. To re-visit the question whether using the (*u', v'*) coordinate improves the correlation between the CCT values and perception, vision experiments were conducted with 12 subjects using a double-booth with non-haploscopic and haploscopic methods. The overall results show that the perception of test lights at CCT 3000 K and Duv -0.015 and 0.015 is closer to the CCT calculated based on CIE 1976 (*u', v'*) coordinate. The perception of CCT at 5500 K with the same Duv shifts appears to be between CCTs based on the (*u, v*) coordinate and the (*u', v'*) coordinate. It is suggested that the definition of CCT be re-examined for possibly using the CIE 1976 (*u', v'*).

Keywords: correlated colour temperature, (*u, v*) diagram, (*u', v'*) diagram; visual perception, white light

1 Introduction

The Correlated Colour Temperature (CCT) is an important quantity for specifying the colour of white light (warm to cool) in lighting applications and others (display and photography). CCT is defined, by International Commission on Illumination (CIE), as "the temperature of the Planckian radiator having the chromaticity nearest the chromaticity associated with the given spectral distribution on a diagram where the (CIE 1931 standard observer based) *u*', 2/3 *v*' coordinates of the Planckian locus and the test stimulus are depicted" (CIE, 2011). This means that the CCT is calculated using the CIE 1960 (*u, v*) diagram (now obsolete), though the current CIE recommendation for uniform colour space is 1976 (*u', v'*) diagram. The CIE 1960 (*u, v*) has been used historically for CCT calculation for decades.

A change to the (*u', v'*) coordinates for CCT calculation was discussed when the CIE 15 (CIE, 2004) was revised in 2004, but no change was made at that time due to lack of research data that would indicate that the (*u', v'*) coordinates would give better correlation with the perception of CCT. There was also no interest from the industry on such a question. However, recently, research on white points below Planckian locus is gaining attention for possibly more natural lighting colour quality (e.g., Ohno and Fein, 2014, Ohno and Oh, 2016) and a fairly large deviation from Planckian locus in negative Duv[1] direction was suggested as white light points perceived most natural. At such white points with large Duv shifts, there would be fairly large differences between CCT values calculated from (*u, v*) and from (*u', v'*). It was also observed during the previous experiments (Ohno and Fein, 2014) that lights having chromaticity below Planckian locus appeared to be at higher CCT. Some users of neodymium lamps (having Duv of approximately -0.005) report the same observation. This shift of perceived CCT is in the same direction as the shift when CCT is calculated based on (*u', v'*).

To specify the position of a white point with respect to Planckian locus, a distance from the Planckian locus on CIE 1960 (u, v) has historically been used because the calculation of CCT has been based on the (u, v) diagram but this term had not been defined in any standard. Duv was defined for the first time in a standard in 2008 (ANSI, 2017). Duv is calculated using CIE 1960 (u, v) coordinates. On the other hand, colour differences and colour shifts of light sources are expressed by the distance $\Delta u'v'$ on (*u', v'*) coordinates as defined by the CIE [CIE, 2014],

[1] Duv, symbol D_{uv}, is the closest distance from the light source chromaticity point to Planckian locus on CIE *u*', 2/3 *v*' coordinates, with + sign for above and - sign for below Planckian locus (ANSI, 2017).

Ha, Hyeyoung; Kwak, Youngshin; Ohno, Yoshihiro.
"Vision Experiment On Perception Of Correlated Colour Temperature."
Paper presented at International Commission on Illumination 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

and there is occasionally confusion in the industry between Duv and $\Delta u'v'$, and question is raised why Duv is based on 1960 (*u, v*). Since Duv is getting widely used, it needs to be defined by CIE but it should be consistent with the CCT definition. Toward such an effort, we determined that the question on the definitions of CCT based on the 1960 (*u, v*) coordinates need to be revisited.

For this purpose, a series of vision experiments have been conducted at National Institute of Standards and Technology (NIST) to investigate which coordinates, (*u, v*), or (*u', v'*), agree more closely with visual perception, when the illumination source chromaticity is deviated from the Planckian locus. The experiments used 12 subjects having normal colour vision and were conducted at a double-lighting booth with spectrally tunable light sources. Two methods, haploscopic and non-haploscopic, were used, as it was not clear which method would work better. The experimental methods and the results are discussed.

2 Experimental Facility

A double-lighting booth with spectrally tunable light sources was used for the experiment. The booth is a commercially available product for visual inspection of products, but the light source part was replaced by the spectrally tunable light sources, which are also commercially available products. The source has 16 channels of light-emitting diode (LED) spectra and provided with a computer control program that allows colour settings by entering CCT and Duv values. The photo of the double-booth is shown in Figure 1. The size of the viewing area on each side is 51 cm wide and 46 cm high, and 64 cm deep.

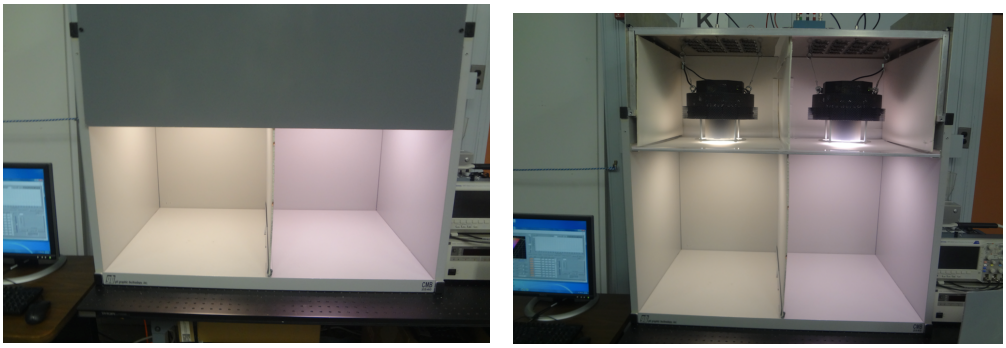


Figure 1 – Photograph of the double-booth used for the experiment. The photos show when the top cover is placed during experiment (left) and when it is removed (right).

The light-emitting surface of the light source is a diffuser with a 10 cm diameter, and there is a large light-transmitting diffuser between the light source compartment and viewing compartment, with which good spatial colour uniformity is provided in the viewing area. It takes a long time (several hours) for the light sources to reach sufficient stability required for this experiment, thus these sources were operated continuously (24 h 7 days) at the reference setting (see section 3.) during the whole experiment period. The spectral distributions (at the centre of the bottom surface) were measured with a spectroradiometer on several light settings before each experiment session (each subject) every day, and the chromaticity drifts during the whole experimental period (about 10 days) was monitored. The drifts were within ± 0.0005 in *u'* or *v'* with a maximum within ± 0.001 from the initial set values and the illuminance was within ± 1.5 % from initial 500 lx setting. The spatial nonuniformity of colour on the booth inner surfaces (bottom surfaces and surrounding wall panels) were within ±0.0008 in *u', v'* from the average value.

Measurements of the booth throughout this experiment was made with an array spectroradiometer in irradiance geometry. The spectroradiometer was calibrated immediately before the experiment period, with a NIST spectral irradiance scale standard lamp. The expanded uncertainty (*k*=2) of the measured chromaticity of the broadband lights used in this experiment was estimated to be within 0.001 in *u', v'*, and the uncertainty of the measurement

Ha, Hyeyoung; Kwak, Youngshin; Ohno, Yoshihiro.
"Vision Experiment On Perception Of Correlated Colour Temperature."
Paper presented at International Commission on Illumination 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

of colour differences (as the case of drifts and spatial uniformity) was estimated to be 0.0002 in *u', v'*. The expanded uncertainly (*k*=2) of the illuminance measurements was 3 % and that of relative changes was 0.2 %.

3 Experimental Settings

The experiments were conducted at two CCT conditions, 3000 K and 5500 K, representing an average for low CCT lamps and an average for high CCT lamps common in the market. A test light, above or below Planckian locus (D_{uv} = +0.015 and $D_{uv=}$ -0.015) was visually compared with the lights on Planckian locus to determine which point on Planckian locus appear closest to the test light. The Duv values are chosen as the values determined as perceived most natural on the average in the previous NIST study [Ohno and Fein, 2014].

The right side of the booth (see Figure 1) was set to a test light, which was at $D_{uv=}$ 0.015 or D_{uv} = -0.015. The left side of the booth was set for nine points on the Planckian locus as shown in Figure 2 (the case for 3000 K). In the figure, point No.5 (labelled "uv") is at exactly the same CCT as the test lights based on the current definition of CCT (calculated based on the 1960 (*u, v*) coordinate), point No. 3, labelled "u'v'(-)", is at the same CCT (3179 K) of the test light (D_{uv} = -0.015) calculated based on the CIE 1976 (*u', v'*) coordinate, and No. 7, labelled "u'v'(+)", is at the same CCT (2882 K) of the test light (D_{uv} = 0.015) calculated based on the (*u', v'*) coordinate. Points No. 4 and 6 are the intermediate points between No. 3 and 5 and between No. 5 and 7, respectively. The intervals of these intermediate points were calculated to be equal from both neighbouring points in Δu'v' measure. Additional two points are extended on each side; No. 1, 2 and No. 8, 9, at the same Δu'v' intervals as the neighbouring intervals. Note that the intervals from No. 1 to 5 and those from No. 5 to 9 are slightly different, but these intervals need not be equal for the experimental methods used (see 4.2 and 4.3).

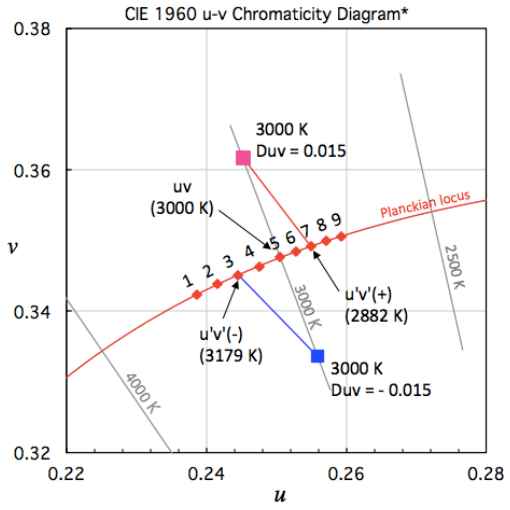


Figure 2 – Chromaticity points set for the double-booth for the experiments (haploscopic) for 3000 K.

The spectral distributions of the sources were set to be as broad as possible using all the 16 channels of the light source, for all these chromaticity points. The spectral data of the lights in all the chromaticity points in Figure 2 (3000 K) and the same set for 5500 K are shown in Figure 3. The calculated chromaticity and colour quality values of all these lights used in the experiment (set values and measured values example) are shown in Table 1. The illuminance for all test lights was set to 500 lx at the bottom surface of the booth. The measurement was made at ≈10 cm above the bottom surface due to the integrating sphere input optics of the spectroradiometer. The illuminance values in Table 1 should be multiplied by 0.85 to convert to the values at the bottom surface of the booth.

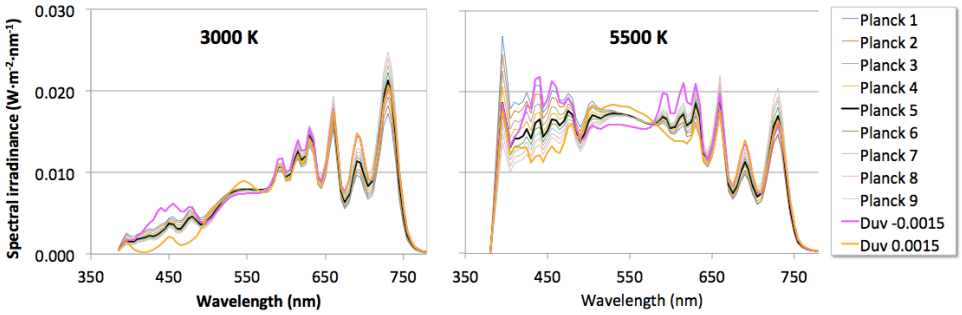


Figure 3 – Spectral power distributions of the lights used in the experiments, for 3000 K (left) and 5500 K (right).

Table 1 – Chromaticity and colour quantity values for the lights used in the experiment.

		Set values						Measured data (an example)				
		CCT	duv	u	v	u'	v'	CCT	CCT dev	duv	Illum. (lx)	CRI (Ra)
Test	Duv=-0.015	3000	-0.015	0.2559	0.3336	0.2559	0.5004	3001	1.4	-0.0153	587.92	92.7
	Duv= 0.015	3000	0.015	0.2452	0.3616	0.2452	0.5424	2998	-2.0	0.0148	587.56	92.0
3000 K Planckian points	1	3384	0.0000	0.2387	0.3424	0.2387	0.5136	3390	6.0	-0.0003	584.34	98.4
	2	3280	0.0000	0.2416	0.3438	0.2416	0.5157	3287	7.2	-0.0003	583.60	98.4
	u'v' - 3	3181	0.0000	0.2446	0.3451	0.2446	0.5177	3186	4.3	0.0003	583.33	98.5
	4	3088	0.0000	0.2476	0.3464	0.2476	0.5195	3093	4.7	0.0003	583.96	98.3
	uv 5	3000	0.0000	0.2506	0.3476	0.2506	0.5214	3006	6.4	0.0002	584.36	98.2
	6	2940	0.0000	0.2527	0.3484	0.2527	0.5226	2945	4.7	0.0002	583.22	98.2
	u'v' + 7	2882	0.0000	0.2549	0.3492	0.2549	0.5238	2887	4.1	0.0001	584.03	98.2
	8	2827	0.0000	0.2571	0.3499	0.2571	0.5249	2832	4.7	0.0001	583.56	98.2
	9	2773	0.0000	0.2593	0.3506	0.2593	0.5259	2777	3.8	0.0002	583.65	98.0
		CCT	duv	u	v	u'	v'	CCT	CCT dev	duv	Illum. (lx)	CRI (Ra)
Test	Duv=-0.015	5500	-0.015	0.2181	0.3084	0.2181	0.4625	5494	-6.3	-0.0154	587.28	90.5
	Duv= 0.015	5500	0.015	0.1957	0.3283	0.1957	0.4925	5506	6.3	0.0149	586.16	91.4
5500 K Planckian points	1	6728	0.0004	0.1993	0.3088	0.1993	0.4632	6722	-6.3	0.0001	585.61	97.9
	2	6375	0.0001	0.2011	0.3113	0.2011	0.4669	6375	-0.7	-0.0002	585.39	97.8
	u'v' - 3	6056	0.0000	0.2030	0.3137	0.2030	0.4705	6055	-1.3	0.0006	585.28	98.0
	4	5765	0.0000	0.2049	0.3160	0.2049	0.4740	5763	-2.1	0.0004	585.06	98.1
	uv 5	5500	0.0000	0.2069	0.3184	0.2069	0.4775	5501	0.7	-0.0003	585.30	98.1
	6	5301	0.0000	0.2086	0.3202	0.2086	0.4802	5305	4.3	0.0004	585.15	98.1
	u'v' + 7	5115	0.0000	0.2103	0.3220	0.2103	0.4830	5118	2.2	0.0003	585.23	98.1
	8	4942	0.0001	0.2120	0.3237	0.2120	0.4856	4945	2.9	0.0004	585.48	99.1
	9	4779	0.0002	0.2138	0.3254	0.2138	0.4881	4783	4.0	0.0004	585.30	99.2

4 Experimental Methods

The experiments were conducted at two CCTs, 3000 K and 5500 K, and at two Duv levels 0.015 and -0.015, therefore, four chromaticity points of the test light. The test light and one of points at Planckian locus were presented side by side. The goal is to determine which Planckian point appear closest to the test light by observation of subjects.

12 subjects having normal colour vision participated in the experiment; 8 males and 4 females, from 20 to 71 years old, 7 white and 5 Asians.

When comparing different colours visually, the adaptation condition of the observer is critical, as perceived colours change depending on adaptation. The adaptation condition must be clearly defined and applied in the experiment. To handle the adaptation conditions appropriately, two methods were used in our experiment. The first method is the non-haploscopic method, in which a subject directly compared different Planckian lights (left side of booth) against the test light (right side). We first tested this method, but we experienced that comparing two fairly different colours of light (one on Planckian, another much off from Planckian) was often difficult. The second method is the haploscopic method, where the subject's left eye and right eye are adapted separately to the light at each side of the booth. With this method, the colours of the

two lights will appear much closer colours due to chromatic adaptation of each eye, thus it was considered that comparisons would be easier.

In non-haploscopic method, subjects looked at the left side of booth (Planckian) and right side (test light with shifted Duv) sequentially by eye movement of both eyes, thus subjects' adaptation was approximately at an average chromaticity of the Planckian and the test light. In the haploscopic method, the subject was adapted to each of Planckian and the test light by separating the view of each eye. This method is considered from the situation in lighting, where observers are always adapted to the illumination and the impression of CCT is given under the adapted condition, thus this method was considered appropriate for lighting situation. As we were not sure which would be the best method, we used both methods under limited resources available.

In addition, in both methods, comparison of colours of the two sides of the booth, when no objects were placed, were found rather difficult, as there is no target to compare for the subjects. We decided to put a large white target sheet at the bottom of the booth so that subjects could compare the colours of the white sheet on both sides, while the surrounding area of the booth was used for subject's adaptation. We did pre-tests with a few subjects, and verified that comparison was easier with the white sheets for the subjects than empty box. This sheet (size, 28 cm x 28 cm) is made of a material based on polytetrafluoroethylene (PTFE) and has very high and spectrally flat reflectance factors (0.96 to 0.97 in 410 nm to 780 nm) as well as near-Lambertian diffuse characteristics. The sheet was also tested to have no fluorescence by excitation around 400 nm. The differences between the chromaticity coordinates of the light source (from spectral irradiance measured at the bottom of the booth) and those of the white target sheet (from spectral radiance measured) are practically negligible (within ±0.0004 in *u', v'*). The spectral reflectance factors of the booth inner surfaces (measured at the center of front wall) was fairly flat, 0.37 to 0.40 in 410 nm to 720 nm. There are slight differences between the chromaticity coordinates of the white target sheet and those of booth inner walls (used for adaptation), but they are within ±0.0016 in *u', v'* and considered insignificant to affect the experimental results.

For haploscopic experiment, a front plate was added to separate the view of each eye of the observer completely. The experimental scenes of the haploscopic and non-haploscopic methods are shown in Figure 4. The subject's eye position was ≈50 cm from the front edges of the booth for non-haploscopic experiment, but it varied to some extent, as a chin rest was not used. In haploscopic experiment, the subject's forehead was contacting the edges of the separating plate mounted in front of the booth having a width of 25 cm. The height of subject eye was at ≈38 to 44 cm from the height of bottom surface of the booth. The height of the chair could be adjusted if the subject was very tall or very short. The cover panel is lower than the height of the diffuser to ensure that light emitting area is shielded from subject's view.

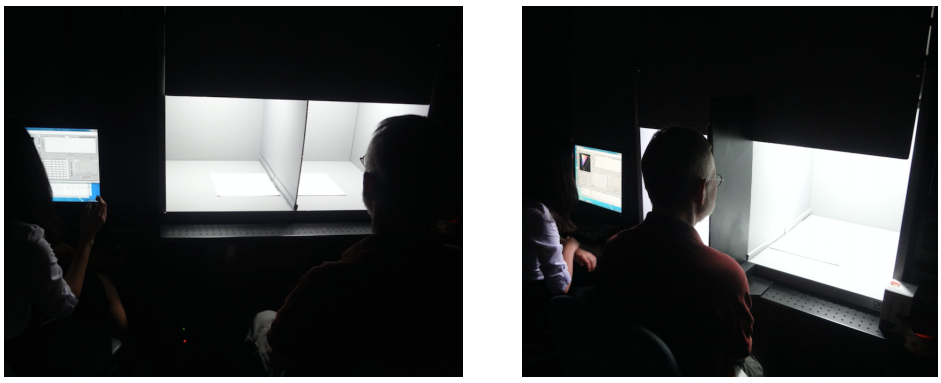


Figure 4 – Experimental scenes of the non-haploscopic method (left) and the haploscopic method (right).

4.1 Initial procedures

These procedures are common with both non-haploscopic and haploscopic methods. When a subject came into the laboratory, he/she was first tested for normal colour vision using Ishihara Test book under room light. Then, room lights were turned off, the subject was adapted for at least 5 min sitting in front of the booth, when both sides of the booth were set for Planckian No. 5 (see Figure 2) of the first CCT (3000 K or 5500 K). Which CCT to start with was randomized for each subject. During this adaptation time, the operator gave introduction of the experiment, instructions for the experiment procedures, then one or two practice runs. Non-haploscopic experiment was always done first, then haploscopic method. Before haploscopic experiment, the subject was asked if their two eyes were not balanced (one eye is dominant). If this is the case (actually there were no such case), the subject would not have participated in the haploscopic experiment.

Due to limitation of experiment time for subjects and limitation of the number of subjects available, the experiments were arranged so that all 12 subjects participated in the haploscopic experiment at both CCT conditions, but they participated in the non-haploscopic experiment at one of CCTs (randomly assigned), thus data for each CCT was obtained for 6 subjects. After all experiments were done, the operator asked the subject which method was easier or any comments, which were also recorded with the results.

4.2 Non-haploscopic experiment procedures

To handle chromatic adaptation appropriately, comparisons were made in pairs of light. In this method, six pairs of Planckian lights were prepared as shown in Figure 5. The Planckian lights are marked number 1 to 9. The light pairs are marked as ① to ⑥ and the adaptation points are green circles between Planckian lights and the test light. A group of comparison experiments for these six pairs is called a "run", and there are four runs; for 3000 K negative Duv, 3000 K positive Duv, 5500 K negative Duv, and 5500 K positive Duv. The order of these four runs was randomized for each subject. Before each run of experiment, the subject was adapted to the first adaptation point (for each CCT and Duv condition) presented on both sides of the booth for 2 min. Then, before comparing each pair, the subject was adapted for 1 min at each adaptation point, which is the meddle point between the test light and the center of the pair of two Planckian points. There are six light pairs for each test light and there are six adaptation points for each light pair. The order of light pairs presented was randomized for each subject.

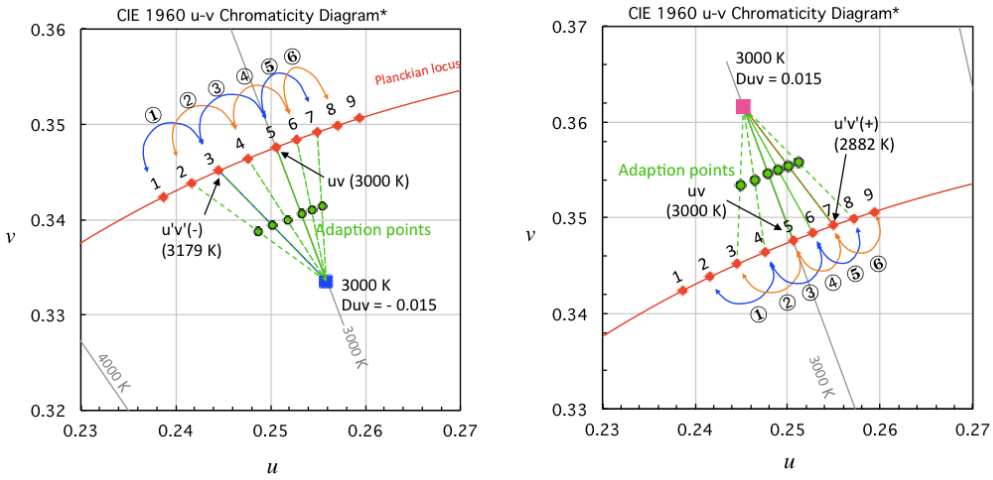


Figure 5 – Test light and pairs of light presented and compared by subjects in haploscopic method, for negative Duv test light (left) and positive Duv test light (right).

For example, in Figure 5 (left), both sides of the booth were first set to the adaptation point between No. 4 and the test light (D_{uv}= -0.015), and the subject was adapted for 2 min. During this time, the subject was instructed to look around the wall areas of booth, and not at the white sheets. Then, the subject was asked to close their eyes momentarily (approximately 10 s) while lights are changed. The left side of booth is switched to Planckian No. 3 and the right side is switched to the test light (D_{uv}= -0.015), then subject was asked to open their eyes. Then, on the left side of the booth, No. 3 and No. 5 were presented alternately for approximately 2 s each time and repeated, while the right side of booth stayed the same (test light). No. 3 is called “A” and No. 5 is called “B”, and the subject is asked which light, A or B, is closer colour to the right side. The order of A and B (for lower or higher number of Planckian light) was randomized. For this comparison, the subject was instructed to compare the colours of the white sheets on the left and right side of the booth. A and B were alternated and repeated several times till the subject could make judgement. This was a forced choice, so the subject was instructed to answer A or B even if the judgment was difficult, but in that case, also tell the operator that judgement was difficult. The number of comment “difficult” would indicate how well the experiments were carried out. The operator recorded these comments together with the answers A or B. The closure of the eyes in the procedure above was necessary because only one side of the booth could be changed at a time, so the two sides had to be changed sequentially, which took several seconds, and occasionally with the light turned off and on, which would disturb subject’s adaptation condition. In this example, if (*u’, v’*) is the correct colour space for CCT, the subject would choose No. 3. If (*u, v*) is the correct colour space for CCT, the subject would choose No. 5.

After completing the first run of experiment, after a short break, next run started with adaptation of 2 min, and another run was performed. Note that only two runs of experiment for one of the CCT conditions were conducted for each subject. Each run took 10 min to 15 min, and the entire time for non-haploscopic experiment including introduction time for each subject was approximately 30 min.

4.3 Haploscopic experiment procedures

In this experiment, the procedures are nearly the same as those for the non-haploscopic experiment described above, except for the adaptation condition for the subjects was different. The experiments used the same nine Planckian chromaticity points, six pairs of light, and two test lights (D_{uv}= -0.015 and D_{uv}= 0.015) as shown in Figure 5 for each CCT condition, but the intermediate adaptation points were not used. Instead, subjects were adapted in a haploscopic condition, with their left eye adapted to the light of left side of the booth (one of Planckian points), and right eye with the light of the right side of the booth (one of the test lights).

Before each run of experiment, the subject was adapted to the first pair of light in haploscopic condition for 2 min. During the initial adaptation time (5 min, see 4.1), a practice run was made under the negative Duv condition, and subject was asked if the colours of left side and right side were becoming closer as his/her eyes are adapted to ensure haploscopic condition is effective for the subject, and also practiced choosing A or B in a light pair.

The order of presenting light pairs was randomized. For example, the left side of booth first presented Planckian No. 4 and the right side of booth presented the test light (D_{uv}= -0.015), and subject was adapted for 1 min in the haploscopic condition. During adaptation time, the subject was instructed to view the wall areas of the booth, and not look at the white sheets. Then, No. 3 is presented as “A” for approximately 2 s, then No. 5 is presented as “B” for 2 s, and these are repeated till the subject had an answer. The transition of the light sources from the adaptation condition to the pair comparison was instant because only the left side of the booth need to be changed, and the subjects did not have to close their eyes in this experiment. The comparisons of lights A and B were conducted in the same manner as described in detail for the non-haploscopic experiment. The comparison of light pairs is repeated for all six pairs for an experiment run.

The subjects had a short break (1 min or so) between the experiment runs with their eyes relaxed in non-haploscopic condition. Each run took 10 min to 15 min, and the entire time for the four runs of experiment for each subject including introduction time was approximately 50 min.

Ha, Hyeyoung; Kwak, Youngshin; Ohno, Yoshihiro.
”Vision Experiment On Perception Of Correlated Colour Temperature.”
Paper presented at International Commission on Illumination 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

5. Results of experiments

Table 2 shows an example of the sorted raw results of four runs of non-haploscopic experiment filled by six subjects. For haploscopic experiment, this set of data was filled by 12 subjects. Light No. 3 is the CCT point calculated with (*u',v'*) diagram for D_{uv}= -0.015 test light, and No.7 is the that for Duv=0.015 test light (the table cells are marked green). No. 5 is the CCT point calculated with (*u, v*) diagram for both D_{uv}= -0.015 and D_{uv}= 0.015 test lights (the table cells are marked yellow). The column "selected" shows which light was selected by the subject for each adaptation point. The cells are coloured blue when higher CCT was chosen and pink when lower CCT was chosen. Therefore, the ideal scenarios for each case, when the perception follows the (*u, v*) coordinate and when the perception follows the (*u', v'*) coordinate, would be the answers as shown in Table 3.

Table 2. An example of the sorted raw results for comparison of six pairs by two subjects.

Subject #	9				10			
Run	1 (3000 K n)		2 (3000 K p)		3 (5500 K n)		4 (5500 K p)	
Test light	Adaptation	Selected	Adaptation	Selected	Adaptation	Selected	Adaptation	Selected
Planckian light No.	2	1	3	4	2	3	3	4
	3	2	4	5	3	4	4	3
	4	3	5	6	4	5	5	4
	5	4	6	5	5	6	6	5
	6	5	7	8	6	5	7	6
	7	6	8	9	7	6	8	7

Table 3. The ideal scenarios of the answers for Table 2.

Subject #	(u',v') scenario				(u,v) scenario			
Run	Negative Duv		Positive Duv		Negative Duv		Positive Duv	
Test light	Adaptation	Selected	Adaptation	Selected	Adaptation	Selected	Adaptation	Selected
Planckian light No.	2	lower CCT	3	lower CCT	2	lower CCT	3	lower CCT
	3	50%	4		3		4	
	4	higher CCT	5		4		5	50%
	5		6		5	50%	6	higher CCT
	6		7	50%	6	higher CCT	7	
	7		8	higher CCT	7		8	

The results for all subjects were sorted in the format of Table 2, and the percentages of higher CCT chosen are calculated for each condition, which are shown in Tables 4 and 5.

Table 4. Average results of the non-haploscopic experiment (n=6)

3000 K (Duv -)		3000 K (Duv +)		5500 K (Duv -)		5500 K (Duv +)	
Adapt.	high CCT %	Adapt.	high CCT %	Adapt.	high CCT %	Adapt.	high CCT %
2	50%	3	0%	2	0%	3	0%
3	67%	4	0%	3	17%	4	33%
4	67%	5	0%	4	50%	5	50%
5	100%	6	17%	5	50%	6	67%
6	83%	7	0%	6	83%	7	50%
7	83%	8	17%	7	83%	8	67%

Table 5. Average results of the haploscopic experiment (n=12)

3000 K (Duv -)		3000 K (Duv +)		5500 K (Duv -)		5500 K (Duv +)	
Adapt.	high CCT %	Adapt.	high CCT %	Adapt.	high CCT %	Adapt.	high CCT %
2	100%	3	0%	2	58%	3	33%
3	100%	4	17%	3	42%	4	42%
4	75%	5	25%	4	75%	5	50%
5	92%	6	25%	5	67%	6	67%
6	100%	7	17%	6	83%	7	33%
7	92%	8	42%	7	83%	8	75%

Ha, Hyeyoung; Kwak, Youngshin; Ohno, Yoshihiro.
"Vision Experiment On Perception Of Correlated Colour Temperature."
Paper presented at International Commission on Illumination 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

The results in Tables 4 and 5 are plotted in Figure 6 and Figure 7. In these figures, the ideal scenarios for the (*u', v'*) coordinate and for the (*u, v*) coordinate, from Table 3, are plotted as dashed lines. Theoretically, the plotted lines of results should be rising from 0 % on low CCT (left) side to 100 % on high CCT (right) side, and the 50% point is considered to be the perceived CCT of the test light.

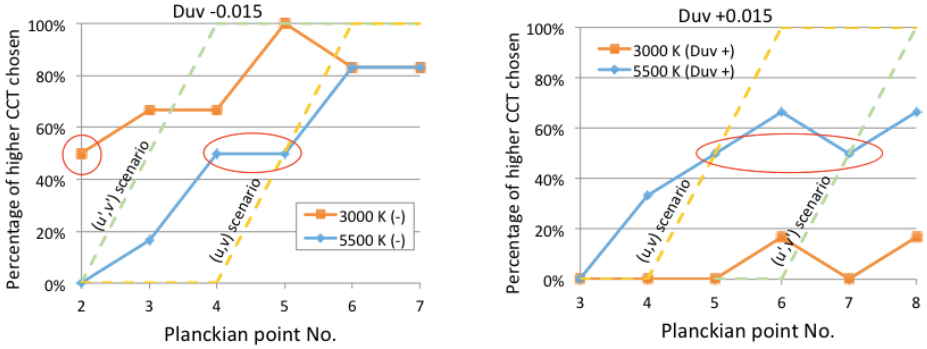


Figure 6 – The results of the non-haploscopic experiment

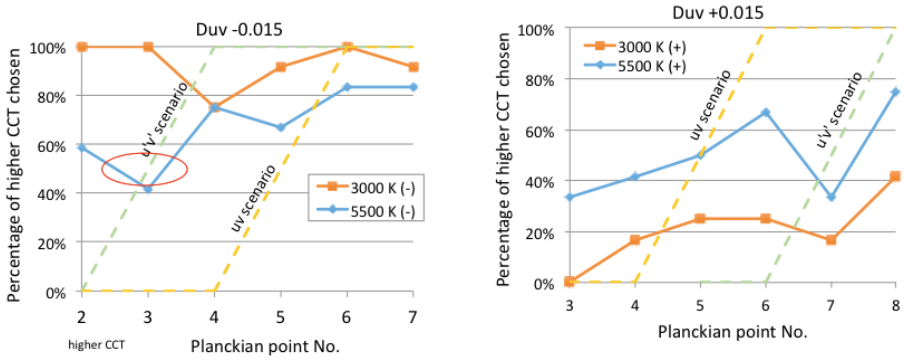


Figure 7 – The results of the haploscopic experiment

From these data shown in the figures, the results are summarized as below:

<Non-haploscopic>

3000 K / D_{uv} -0.015	perceived CCT is closer to the u'v' point.
3000 K / D_{uv} 0.015	perceived CCT is closer to the u'v' point (beyond).
5500 K / D_{uv} -0.015	perceived CCT is closer to uv point.
5500 K / D_{uv} 0.015	perceived CCT is between the uv point and u'v' point.

<Haploscopic>

3000 K / D_{uv} -0.015	perceived CCT is closer to the u'v' point (beyond).
3000 K / D_{uv} 0.015	perceived CCT is closer to the u'v' point (beyond).
5500 K / D_{uv} -0.015	perceived CCT is closer to the u'v' point.
5500 K / D_{uv} 0.015	perceived CCT is between the uv point and u'v' point.

The results above are further summarized that the perceived CCT at 3000 K is consistently closer to the u'v' point (or beyond), and the perceived CCT at 5500 K is between the uv point and u'v' point. The two methods produced fairly consistent results between them. These results may be explained by observation of MacAdam ellipses (e.g., see Figure 3 of CIE, 2014). The ellipses are close to a circle at 3000 K on the (*u', v'*) diagram, but they are closer to a circle on the (*u, v*) diagram at 6500 K. The results above are consistent with this observation on MacAdam ellipses.

Both methods worked fairly well. But for comparison, judging from the rate of “difficult to judge”, it seems that non-haploscopic method was more difficult (16 %) than haploscopic method (9 %). Comparing CCT conditions, 5500 K condition was more difficult (14 %) than 3000 K (8 %). Comparing positive and negative Duv, it appears that negative Duv was more difficult (14 %) than positive Duv (9 %). However, these differences are not so significant.

Conclusions

Vision experiments for perception of correlated colour temperature were conducted with 12 subjects. The results in this study indicate that the perception of CCT at 3000 K is closer to the CCT calculated based on CIE 1976 (*u', v'*) coordinate and possibly with even further shifts. The perception of CCT at 5500 K is shown as between the (*u, v*) coordinate and the (*u', v'*) coordinate and is less conclusive. The results obtained agree with general experience in lighting that negative Duv lights, especially those at lower CCTs (e.g., neodymium lamp), appear higher CCT than currently specified values based on (*u, v*) coordinate. While further experiments are desired with more number of subjects and/or by different laboratories, it is suggested that the definition of CCT be re-examined for using the CIE 1976 (*u', v'*) coordinate for the calculation and possibly reforming the definition of Duv accordingly.

Acknowledgements

The authors thank Steve Paolini of Telelumen, LLC for improving the control program for the light sources used in this experiment, on our requests. This research was conducted when Y. Kwak and H. Ha stayed at NIST as guest researchers.

References

ANSI 2017. ANSI C78.377-2017. Specifications of the Chromaticity for Solid State Lighting Products (The first version was published in 2008).

CIE 2004. CIE 15: 2004. Colourimetry, 3rd ed.

CIE 2011. CIE S 017/E:2011. ILV: International Lighting Vocabulary.

CIE 2014. CIE TN 001. Chromaticity Difference Specification for Light Sources.

Ohno, Y., Fein, M., 2014. Vision Experiment on Acceptable and Preferred White Light Chromaticity for Lighting, CIE x039:2014, pp. 192-199.

Ohno, Y., Oh, S., 2016. Vision Experiment II on White Light Chromaticity for Lighting, Proc., CIE 2016 LQ&EE, Melbourne, CIE x042:2016, pp. 175-184.

VISION EXPERIMENT ON CHROMA SATURATION PREFERENCE IN DIFFERENT HUES

OHNO, Y.[1], OH, S.[2], KWAK, Y.[2]
[1] National Institute of Standards and Technology, Gaithersburg, USA
[2] Ulsan National Institute of Science and Technology, Republic of Korea
ohno@nist.gov

Abstract

Increase of chroma is known to be a major factor for colour quality preference in lighting, and gamut area measures are often used to evaluate preference aspects. However, gamut area accounts for chroma differences equally for all hue directions. Experiences in lighting indicates that red is most dominant in such colour quality preference but experimental data has not been available. Vision experiments were conducted to evaluate different effects of chroma shifts in different hue directions on colour quality preference, using NIST Spectrally Tunable Lighting Facility simulating an interior room. 19 subjects viewed various fruits, vegetables, and their skin tones, under illumination of 11 different spectra of different gamut shapes. The 11 spectra were presented as pairs in all combinations, at correlated colour temperatures (CCT) of 2700 K, 3500 K, 5000 K, (D_{uv}= 0) and 3500 K (D_{uv}= -0.015). The average results show that chroma increases of red (ΔC^*_{ab} ≈ 5 to 15) has the largest effect, green (ΔC^*_{ab} ≈15) has the second effect, and yellow's effect is insignificant, in subjects' preference. The results will be useful to develop colour preference metrics that correlates well with perceived colour quality.

Keywords: colour preference, colour quality, chroma, hue, gamut area, lighting

1 Introduction

The overall colour quality of lighting as experienced by users is affected not only by colour fidelity such as International Commission on Illumination (CIE) Colour Rendering Index (CRI), but also other perception effects beyond colour fidelity, especially colour quality preference. Increase of chroma is known to be a major factor for colour quality preference in lighting, and gamut area metrics have been proposed as a preference measure. Gamut area metrics recently proposed include Feeling of Contrast Index (Hashimoto et al, 2007), Gamut Area Index (GAI) (Rea and Freyssinier, 2010), gamut area Q_g in Colour Quality Scale (CQS) (Davis and Ohno, 2010), and most recently, Gamut Index R_g in Illuminating Engineering Society (IES) TM-30 (IES, 2015). However, gamut area accounts for chroma differences equally in all directions of hue. Experiences in lighting indicate that preference effects depend much on hue of object colour, in particular, red is considered to be dominant in colour quality preference, but such experimental data for lighting have not been available.

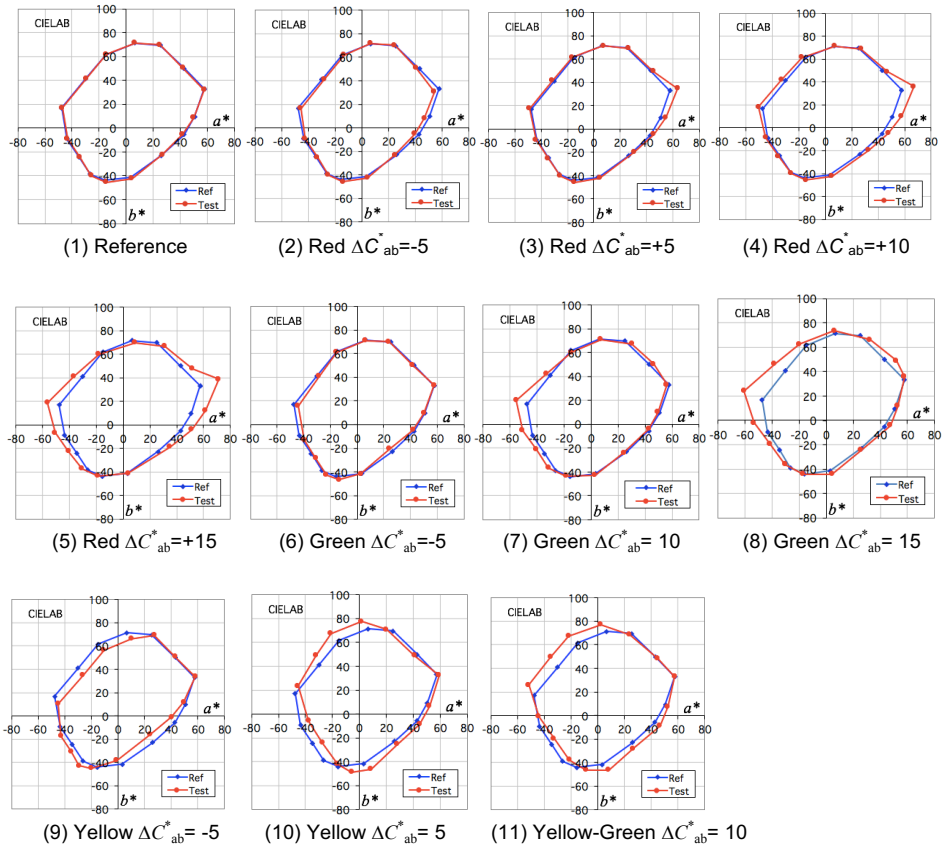
Following the experiment on chroma saturation at National Institute of Standards and Technology (NIST) (Ohno et al, 2015), a series of vision experiments have been conducted to compare the effects of chroma increase or decrease in different hues on colour quality preference, using NIST Spectrally Tunable Lighting Facility (STLF) (Miller et al, 2009) simulating an interior room. 19 subjects participated in the experiments and evaluated their preference on colour quality of objects on the table, which were various real fruits and vegetables of mixed colours, and their skin tones also, under illumination of 11 different spectra of different gamut shapes presented as pairs in all combinations. The experiments were conducted at correlated colour temperatures (CCT) of 2700 K, 3500 K, 5000 K (each D_{uv}=0), and at a negative Duv [1] condition (D_{uv}= -0.015) at 3500 K. The experimental procedures and results are reported.

[1] Duv, symbol D_{uv}, is the closest distance from the light source chromaticity point to Planckian locus on CIE *u*', 2/3 *v*' coordinates, with + sign for above and - sign for below Planckian locus (ANSI, 2017).

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

2 Experimental settings with NIST STLF

For the experiment at each CCT/Duv condition, 11 different chroma saturation conditions as shown in Figure 1 were prepared, which had chroma increase or decrease in red, green, and yellow directions, made as independent as possible from other colors. The figure shows the CIELAB (a^*, b^*) plots of 15 saturated Munsell samples as used in Colour Quality Scale (CQS) (Davis and Ohno, 2010). The red, green, yellow, and yellow-green samples used for chroma settings were 5R4/14, 2.5G6/12, 5Y8/12, and 7.5GY7/10, respectively. Figure 1 shows only data for 3500 K (D_{uv}=0) condition, and the same set of 11 lights were prepared also for 2700 K, 5000 K (both D_{uv}=0), and 3500K (D_{uv}= -0.015), thus total 44 lights were prepared. In the figure, the red lines show the gamut of the measured test light, and the blue lines show the gamut of the reference illuminant, which is the same as that used for Colour Rendering Index (CRI) calculation (Planckian radiation or a phase of daylight of the same CCT as the test light). The ΔC^*_{ab} values are shifts of the chroma from the reference illuminant. Positive and negative numbers mean increase or decrease, respectively, of chroma from the reference illuminant. The ΔC^*_{ab} values are chroma shifts from the reference illuminant to the test light, and these values are target values; measured values were slightly different. The spectra of STLF lights were tuned so that, when setting chroma shifts of one colour (e.g., green), the chroma of other colours (e.g., red) was kept as close as possible to ΔC^*_{ab} =0 (the point for the reference) so that the saturation effect of only one colour could be evaluated, though this was difficult and not always possible. The illuminance was set at ≈250 lx ±1 % on the table in STLF for all light conditions.



(1) Reference (2) Red ΔC^*_{ab}=-5 (3) Red ΔC^*_{ab}=+5 (4) Red ΔC^*_{ab}=+10

(5) Red ΔC^*_{ab}=+15 (6) Green ΔC^*_{ab}=-5 (7) Green ΔC^*_{ab}= 10 (8) Green ΔC^*_{ab}= 15

(9) Yellow ΔC^*_{ab}= -5 (10) Yellow ΔC^*_{ab}= 5 (11) Yellow-Green ΔC^*_{ab}= 10

Figure 1– Gamut shapes on CIE (a^*, b^*) plane of the 11 spectra at 3500 K (D_{uv}=0) used for the experiment. Blue curves show the (a^*, b^*) plots of the 15 saturated Munsell samples for the reference illuminant, and the red curves show those for the 11 test lights.

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

It was difficult to control the chroma in blue direction independently from chroma changes in other colours, and experiment for blue could not be implemented. There were also large hue shifts in blue when controlling blue chroma. It is also considered that increase or decrease in blue is unlikely with the actual lighting products having reasonable colour rendering.

All channels of STLF were utilized to produce these chroma settings, except the 405 nm channel, to avoid any effects of fluorescence from viewed objects. The spectral distributions of all 11 lights for all CCT/Duv conditions are shown in Figure 2. The reference lights had CRI R_a values of ≈ 98 except the D_{uv}= -0.015 condition where R_a value of the reference was 93. For other test lights, R_a values ranged from 63 to 95. ΔC^*_{ab} values were set mostly within ±1 unit from the target value, with a few exceptions of up to ±2 units from the target values.

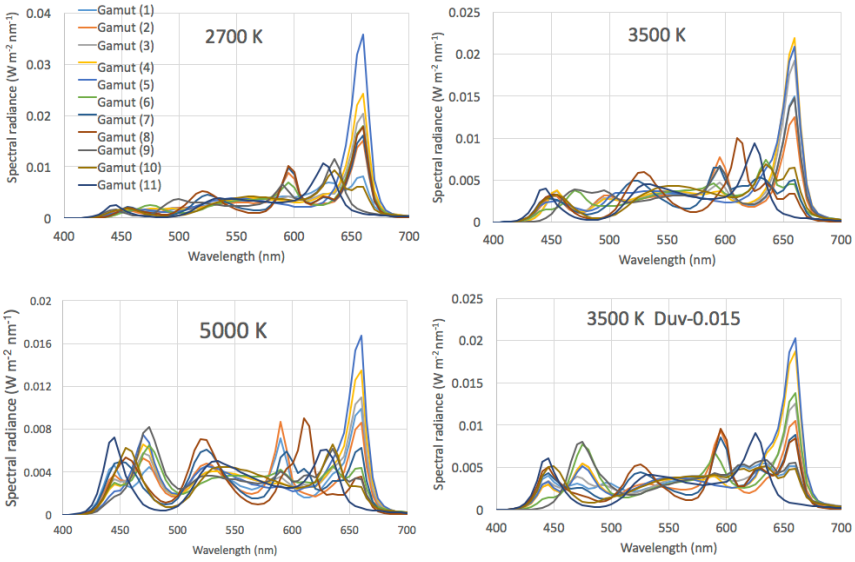


Figure 2 – The spectral distributions of the 11 lights used in the experiment at each CCT/Duv conditions

The STLF cubicle for the experiment was arranged as an interior room as shown in Figure 3, with a couch, a coffee table, a bookshelf with books, some artificial flowers, paintings on the walls. A mirror was also placed on the wall against the couch, to allow evaluation of skin tone of the face of the subject. On the table, there were two plates of real fruits and vegetables; apples, oranges, bananas, strawberries, green peppers, lettuce, tomatoes, red cabbage, and grapes (green and red).

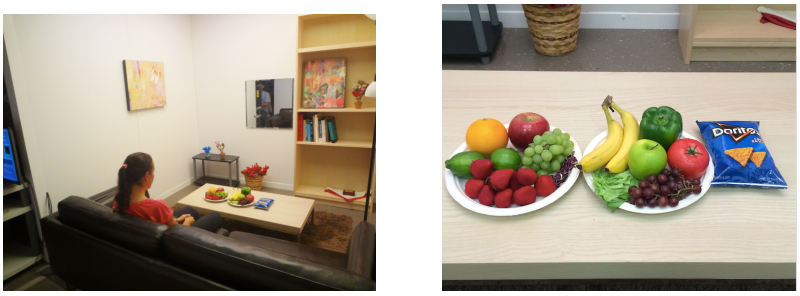


Figure 3 – The setup of the STLF cubicle (left), the fruits and vegetables on the table, used in the experiment

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

These fruits and vegetables were replaced every few days to keep them fresh during the experiment period of about three weeks. Pictures of these fruits and vegetables were taken, and when these were replaced, those having as similar colours and sizes as possible were purchased and used throughout the experiments.

3 Experimental Procedures

19 subjects having normal colour vision participated in the experiment; 11 males and 8 females, from 18 to 84 years old, 13 white and 6 Asians. Each subject evaluated their preference on colour appearance of the target objects under illumination of the 11 different spectra shown above as pairs of lights in all combinations (55 pairs total), which formed an experimental session. Therefore, each of 11 test lights was compared with all other 10 test lights in pairs under the same CCT and Duv condition. Five experimental sessions for different CCT/Duv conditions were conducted for each subject as shown in Table 1. The experiments on their skin tones (subject's hands and face in a mirror) were conducted only at 3500 K due to limitation of time. The experimental session at 3500 K for fruits and vegetables was repeated to test reproducibility. Thus, these five experimental sessions were conducted for each subject, requiring approximately two and half hours in total for each subject.

Table 1 – Conditions for the 5 experiment-runs for each subject

Session	CCT [K]	Duv	Viewing target
1	3500	0	Fruits and vegetables
2	2700	0	Fruits and vegetables
3	5000	0	Fruits and vegetables
4	3500	-0.015	Fruits and vegetables
5	3500	0	Skin tone

Each subject was first tested for normal colour vision using Ishihara Test. Then, instructions for experiment were given, with a few trial comparisons of pairs. The subject was instructed to pay attention to all of red, green, and yellow fruits and vegetables (not look at only one object) when comparing a pair of lights. These initial procedures were done at 3500 K setting, requiring about 10 min, which also served as the first adaptation for the subject. When CCT is changed, the subject was adapted for three min before pair comparisons started. The 55 pairs were presented in random order. Each pair of light was presented sequentially and repeatedly as "A" and "B", for two or three seconds each, and subjects answered which light, A or B, he or she preferred (forced choice), when viewing targets in the room. The order of the test light and reference light in each pair was also randomized. While the answer was forced choice, the subject was also instructed to say, "it is difficult" if the choice was difficult and it was recorded.

Typically, two subjects per day were scheduled for experiment. The STLF was stabilized for at least 30 min before experiments started each day. Before and after all the experiments for each day, the chromaticity (CCT and Duv) of all 44 lights were measured and recorded to verify that the STLF was working in stable conditions. The spectral power distribution of all 44 lights were measured before and after the entire experiment period, to verify that the light spectra were stable during the experiment period. These stability checks ensured that the SLTF for all settings reproduced and was stable to within ±0.001 in *u', v'* during the experiment period.

4 Results

From the raw results of 55 pair comparisons for 19 subjects, first the proportions that each light (No. 1 to No. 11) was chosen in comparison to other 10 lights in each CCT/Duv condition were calculated. **Table 2** shows such results for 3500 K as an example. No. 1 to 11 in the first column and the first row correspond to the 11 lights of different gamut shapes shown in Figure 1. The values in parenthesis are ΔC^*_{ab} values of chroma shift from the reference. For example, in a pair of Reference (0) and Red (5), 89% (17 out of 19) of subjects preferred Red (ΔC^*_{ab}= 5) to Red (ΔC^*_{ab}= 0). Similarly, in a pair of Reference (0) and Red (-5), only 11% (2 out of 19) of subjects preferred Red (ΔC^*_{ab}=-5) to Red (ΔC^*_{ab}= 0).

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

Table 2 – Proportions of light chosen in 55 pairs (3500 K)

	No. (i)	1	2	3	4	5	6	7	8	9	10	11
No. (j)		Ref (0)	Red (-5)	Red (5)	Red (10)	Red (15)	Green (-5)	Green (10)	Green (15)	Yellow (-5)	Yellow (5)	YG (10)
1	Ref (0)	0.50	0.11	0.89	0.89	0.74	0.26	0.63	0.84	0.42	0.53	0.68
2	Red (-5)	0.89	0.50	0.84	0.89	0.68	0.63	0.95	0.84	0.58	0.84	0.84
3	Red (5)	0.11	0.16	0.50	0.79	0.63	0.21	0.37	0.53	0.16	0.32	0.26
4	Red (10)	0.11	0.11	0.21	0.50	0.58	0.11	0.32	0.47	0.16	0.21	0.32
5	Red (15)	0.26	0.32	0.37	0.42	0.50	0.21	0.26	0.37	0.21	0.32	0.37
6	Green (-5)	0.74	0.37	0.79	0.89	0.79	0.50	0.68	0.84	0.42	0.68	0.53
7	Green (10)	0.37	0.05	0.63	0.68	0.74	0.32	0.50	0.84	0.37	0.53	0.47
8	Green (15)	0.16	0.16	0.47	0.53	0.63	0.16	0.16	0.50	0.16	0.21	0.21
9	Yellow (-5)	0.58	0.42	0.84	0.84	0.79	0.58	0.63	0.84	0.50	0.58	0.47
10	Yellow (5)	0.47	0.16	0.68	0.79	0.68	0.32	0.47	0.79	0.42	0.50	0.42
11	YG (8)	0.32	0.16	0.74	0.68	0.63	0.47	0.53	0.79	0.53	0.58	0.50

The data in Table 2 were then converted to z-scores, which are often used in psychophysics data analyses (Gescheider 2013). The z-score converts probability data into the scale of the evaluated quantity, which is subjects' preference of lighting in this case. The values in Table 2 are taken as probability $p_{i,j}$ that light i was chosen as preferred over light j. The z-scores $z_{i,j}$ for each pair of light i vs light j were calculated as follows.

A probability density function of the standard normal distribution (normal distribution with its mean equal to 0 and its standard deviation equal to 1) is considered:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{1}$$

The integrated value of this function from minus infinity to the value x is calculated, and $z_{i,j}$ is determined as the value of x when the integral is equal to $p_{i,j}$, as

$$z_{ij} = x; \quad \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = p_{ij} \tag{2}$$

The calculated z-score values for 3500K condition are shown in Table 3 as an example. These values indicate the degree of preference of light i in the scale from negative (disliked) to positive (liked) with respect to light j. The average scores, $z_{ave,i}$, for each light i are calculated, then these are normalized by the values for the reference light, $z_{norm,i}$, as shown at the bottom two rows of Table 3. The normalized z-scores show the preference level of each light compared to the reference light (neutral saturation). Positive numbers mean preferred, negative numbers mean disliked, compared to the reference light, and their absolute values show the degrees of being preferred or disliked. Table 4 shows the average z-score values and Table 5 for the normalized z-score values for all other CCT/Duv conditions.

Table 3 – z-scores for all 55 pair comparisons (3500 K)

	No. (i)	1	2	3	4	5	6	7	8	9	10	11
No. (j)		Ref (0)	Red (-5)	Red (5)	Red (10)	Red (15)	Green (-5)	Green (10)	Green (15)	Yellow (-5)	Yellow (5)	YG (10)
1	Ref (0)	0.00	-1.25	1.25	1.25	0.63	-0.63	0.34	1.00	-0.20	0.07	0.48
2	Red (-5)	1.25	0.00	1.00	1.25	0.48	0.34	1.62	1.00	0.20	1.00	1.00
3	Red (5)	-1.25	-1.00	0.00	0.80	0.34	-0.80	-0.34	0.07	-1.00	-0.48	-0.63
4	Red (10)	-1.25	-1.25	-0.80	0.00	0.20	-1.25	-0.48	-0.07	-1.00	-0.80	-0.48
5	Red (15)	-0.63	-0.48	-0.34	-0.20	0.00	-0.80	-0.63	-0.34	-0.80	-0.48	-0.34
6	Green (-5)	0.63	-0.34	0.80	1.25	0.80	0.00	0.48	1.00	-0.20	0.48	0.07
7	Green (10)	-0.34	-1.62	0.34	0.48	0.63	-0.48	0.00	1.00	-0.34	0.07	-0.07
8	Green (15)	-1.00	-1.00	-0.07	0.07	0.34	-1.00	-1.00	0.00	-1.00	-0.80	-0.80
9	Yellow (-5)	0.20	-0.20	1.00	1.00	0.80	0.20	0.34	1.00	0.00	0.20	-0.07
10	Yellow (5)	-0.07	-1.00	0.48	0.80	0.48	-0.48	-0.07	0.80	-0.20	0.00	-0.20
11	YG (8)	-0.48	-1.00	0.63	0.48	0.34	-0.07	0.07	0.80	0.07	0.20	0.00
	$Z_{ave,i}$	-0.27	-0.83	0.39	0.65	0.46	-0.45	0.03	0.57	-0.41	-0.05	-0.09
	$Z_{norm,i}$	**0.00**	**-0.56**	**0.66**	**0.92**	**0.73**	**-0.19**	**0.30**	**0.84**	**-0.14**	**0.22**	**0.17**

Table 4 – The average z-score values, $z_{ave,i}$, for the 11 lights for all CCT/Duv conditions

	Ref (0)	Red (-5)	Red (5)	Red (10)	Red (15)	Green (-5)	Green (10)	Green (15)	Yellow (-5)	Yellow (5)	YG (10)
2700K	-0.05	-0.81	0.44	0.39	0.13	-0.37	-0.22	0.25	-0.18	0.22	0.20
3500K	-0.27	-0.83	0.39	0.65	0.46	-0.45	0.03	0.57	-0.41	-0.05	-0.09
5000K	-0.09	-0.84	0.40	0.73	0.66	-0.66	-0.10	0.37	0.11	-0.24	-0.34
3500K, -Duv	-0.26	-0.82	0.45	0.54	0.46	-0.08	0.10	0.08	-0.15	-0.24	-0.09
3500K (Skin)	-0.10	-0.59	0.50	0.46	0.21	0.07	-0.21	0.25	0.10	-0.26	-0.44
All CCT/Duv	**-0.15**	**-0.78**	**0.44**	**0.55**	**0.39**	**-0.30**	**-0.08**	**0.31**	**-0.10**	**-0.11**	**-0.15**

Table 5 – The normalized z-score values, $z_{norm,i}$, for the 11 lights for all CCT/Duv conditions

	Ref (0)	Red (-5)	Red (5)	Red (10)	Red (15)	Green (-5)	Green (10)	Green (15)	Yellow (-5)	Yellow (5)	YG (10)
2700K	0.00	-0.76	0.49	0.44	0.19	-0.31	-0.16	0.31	-0.12	0.28	0.25
3500K	0.00	-0.56	0.66	0.92	0.73	-0.19	0.30	0.84	-0.14	0.22	0.17
5000K	0.00	-0.75	0.50	0.82	0.75	-0.57	-0.01	0.46	0.20	-0.15	-0.25
3500K, -Duv	0.00	-0.56	0.71	0.79	0.72	0.17	0.36	0.34	0.11	0.02	0.17
3500K (Skin)	0.00	-0.50	0.60	0.55	0.31	0.17	-0.12	0.35	0.20	-0.17	-0.34
All CCT/Duv	**0.00**	**-0.63**	**0.59**	**0.71**	**0.54**	**-0.15**	**0.07**	**0.46**	**0.05**	**0.04**	**0.00**

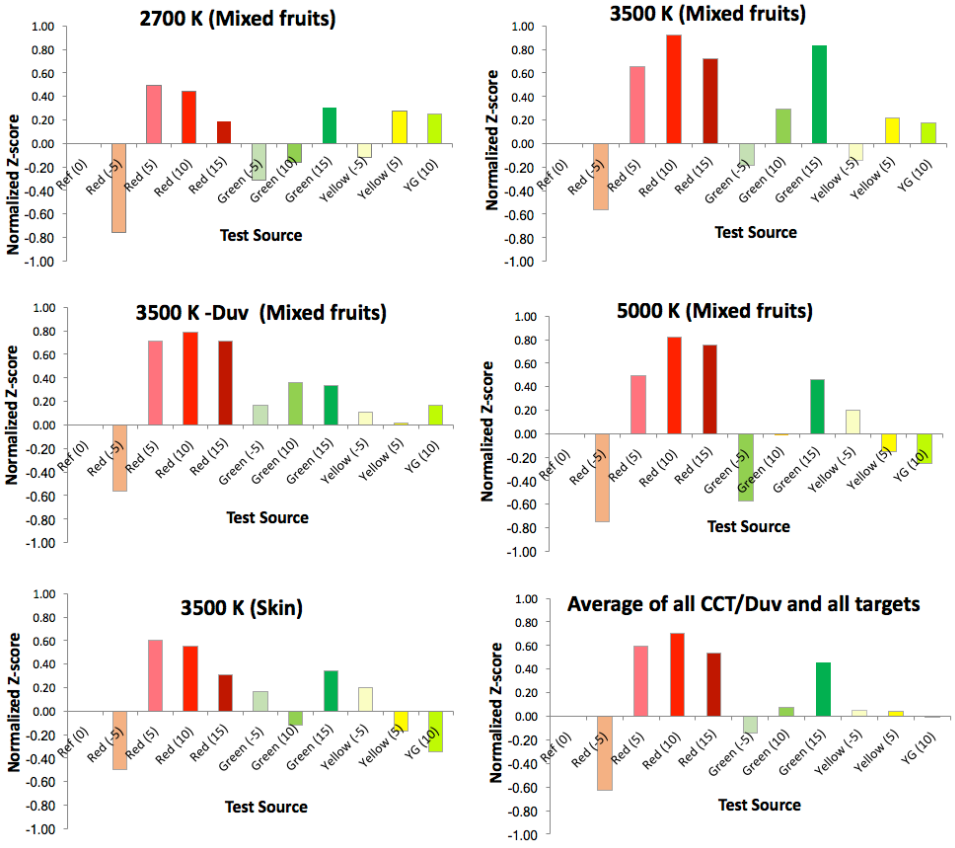


Figure 4 – Preference effects for chroma shift in different hues, for different CCT/Duv conditions and viewing targets.

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

Figure 4 shows all the results in Table 5 in a graphic form. While there are some variations for different CCT/Duv conditions, the results overall show that red has the dominant effect, and preference is sensitive to small increase or decrease of red chroma. Green is the second, however, preference for green chroma increase is less sensitive than red and it requires large chroma increase to be effective. Yellow and yellow-green show much less effect than red or green. In the 3500 K skin tone condition, green (15) shows high preference. It is considered that green should not affect much the preference of skin tone. It is observed that the gamut curve for Green ΔC^*_{ab}= 15 in Figure 1 shows slight increase of chroma in the red region, though the particular red sample's chroma is kept nearly equal to that of the reference. It is considered that this slight increase of chroma around the red region may have affected the results for the skin tone, and also for the green (15) results for fruits and vegetables, because red has strong effects.

Figure 5 shows the plots of the normalized z-score values from the same data in Table 5, arranged for each colour (hue), presented as a function of chroma shift. The curves for Red (Mixed fruits) and Red (Skin) have peaks around ΔC^*_{ab}= 5 to ΔC^*_{ab}= 10, which are very similar to the results in our previous study (Ohno et al, 2015). The curves for Green (Mixed fruits) show much less slopes than those for Red. The curves for yellow show varied slopes depending on CCT and the effects are not clear. These observations support earlier conclusion that red is clearly dominant in colour quality preference and there seem to be optimum levels of red chroma increase for high preference.

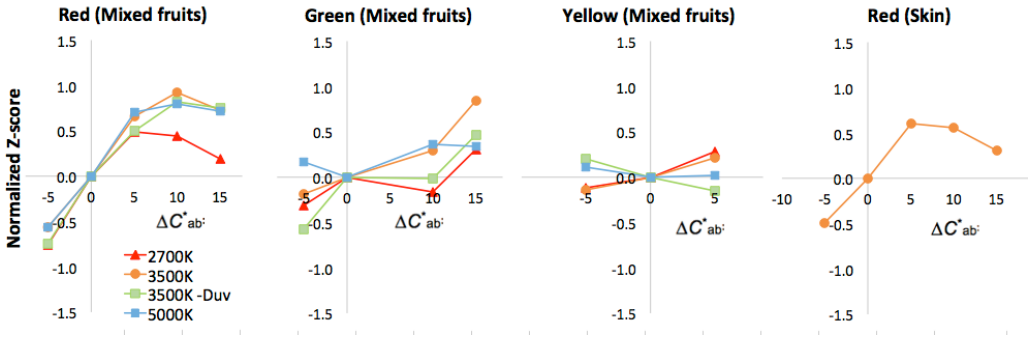


Figure 5 – Preference effects as a function of chroma shift, for different hues and targets.

5 Conclusions

A vision experiment has been conducted using 19 subjects, for the preference effect of chroma shifts in different hue directions, for red, yellow, yellow-green, and green. The chroma of other colors was kept close to the reference. The results proved that the chroma increase of red colour (hue) has the dominant effect on colour preference among these colours, and the results for red were consistent with the previous study (Ohno et al, 2015). The effect of green saturation is the second, but its effect is much less sensitive than red. The effects of yellow and yellow-green are found inconsistent and much less sensitive than red chroma shift.

These results suggest that, for evaluating colour preference of a light source, chroma shifts in different hues by the light source should be weighted differently, with highest weight for red, and much less weight for other colours. Further experimental data will be needed to develop an accurate colour preference model. In this experiment, though it was hoped to test all hues, only red – yellow – green region could be covered due to the difficulty in spectral control. Blue chroma could not be controlled independently from other colors. Yellow shifts also accompany large hue shifts and it was difficult to control yellow chroma independently. These imperfections of light settings may have affected some results. Further experiments are desired with more accurate independent control of chroma for a larger range of hue, e.g., by using a display simulating real objects as used in this study.

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

Acknowledgement

The authors thank Dr. Yuki Kawashima at NIST for his valuable discussions on the data analysis using *z*-score. This research was conducted when Youngshin Kwak and Semin Oh stayed as guest researchers at NIST.

References

ANSI 2017. ANSI C78.377-2017 Specifications for the Chromaticity of Solid State Lighting Products.

Davis, W. and Ohno, Y. 2010. Color Quality Scale, *Optical Engineering*, 033602 March 2010/Vol. 49 (3).

Gescheider, G. A. 2013. Psychophysics: The fundamentals, 3rd ed. Lawrence Eribaum Associates, Mahwah, New Jersey

Hashimoto, K., Yano, T., Shimizu, M., and Nayatani, Y. 2007. New method for specifying colour-rendering properties of light sources based on feeling of contrast," Colour Res. Appl. 32, 361–371.

IES 2015. IES TM-30-15 IES Method for Evaluating Light Source Colour Rendition, Illuminating Engineering Society of North America, New York, NY.

Miller, C., Ohno, Y., Davis, W., Zong, Y., Dowling, K. 2009. "NIST spectrally tunable lighting facility for colour rendering and lighting experiments," in Proc. CIE 2009: Light and Lighting Conference. 5 pages (2009).

Ohno, Y., Fein, M., Miller, C. 2015. Vision Experiment on Chroma Saturation for Color Quality Preference, Proc. 28th Session of CIE, CIE 216: 2015, pp. 60-69 (2015).

Rea, M., Freyssinier, J. 2010. Colour rendering: Beyond pride and prejudice, Colour Res. Appl. 35, Issue 6.

Kwak, Youngshin; Oh, Semin; Ohno, Yoshihiro.
"Vision Experiment on Chroma Saturation Preference in Different Hues."
Paper presented at International Commission on Illumination (CIE) 2017 Midterm Meeting, Jeju, Republic of Korea. October 23, 2017 - October 25, 2017.

Proceedings of the ASME 2017 International Mechanical Engineering Congress and Exposition
IMECE2017
November 3-9, 2017, Tampa, Florida, USA

IMECE2017-70899

Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums

Craig M. Shakarji
Physical Measurement Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.
craig.shakarji@nist.gov

Vijay Srinivasan
Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.
vijay.srinivasan@nist.gov

Abstract

This paper addresses some important theoretical issues for constrained least-squares fitting of planes and parallel planes to a set of input points. In particular, it addresses the convexity of the objective function and the combinatorial characterizations of the optimality conditions. These problems arise in establishing planar datums and systems of planar datums in digital manufacturing. It is shown that even when the input points are in general position: (1) a primary planar datum can contact 1, 2, or 3 input points, (2) a secondary planar datum can contact 1 or 2 input points, and (3) two parallel planes can each contact 1, 2, or 3 input points, but there are some constraints to these combinatorial counts. In addition, it is shown that the objective functions are convex over the domains of interest. The optimality conditions and convexity of objective functions proved in this paper will enable one to verify whether a given solution is a feasible solution, and to design efficient algorithms to find the global optimum solution.

1. Introduction

Planar datums and systems of planar datums arise frequently in the specification and verification of product geometries before, during, and after manufacturing [1-7]. Traditionally, such datum planes were established on a manufactured part using physical devices such as surface plates, angle blocks, and expanding and closing vises [8]. In the current era of digital manufacturing, one faces the task of establishing datums by *fitting* planes and lines to a cloud of points, which may number in the millions, sampled on a manufactured part using coordinate measuring systems (CMS). Standards development organizations such as ISO (International Organization for Standardization) and ASME are now responding to this trend by moving beyond supporting merely analog (i.e., physical) inspection devices to more general standards definitions that also support such digital (i.e., coordinate measurement) technologies. Experts in ISO/TC 213 and ASME Y14 standards committees are now engaged in defining the proper fitting criteria that simulate in the digital world what has been practiced in the physical world thus far.

Mathematically and computationally, fitting can be viewed as an optimization problem. Least-squares (including total least-squares) fitting of lines and planes has a long and colorful history over the past two centuries in many fields of science and engineering [9-11]. In computational coordinate metrology, least-squares fitting has enjoyed an enduring appeal [12, 13] that got a boost from recent interest in the form of constrained least-squares fitting to establish datums. Fueled by urgent requests from ISO and ASME standards committees, recent research has explored some of the theoretical and algorithmic issues of constrained least-squares fitting [14, 15]. These investigations were directed towards planar datums and systems of planar datums, which influenced the standards committees to consider constrained least-squares fitting as the default datum fitting criterion.

This paper consolidates earlier theoretical results for constrained least-squares fitting of planes, completes them with formal proofs, and extends the results to parallel planes and intersecting planes. It will address only the combinatorial characterizations of the optimality conditions and the convexity of objective functions, leaving the algorithmic details to other published sources and future research. However, the theoretical results reported this paper form the basis for much of the algorithm design and analysis.

The rest of the paper is organized as follows. Section 2 provides some motivating examples for planar datums. The optimization problem for plane fitting is then formulated in Section 3. A convex hull filter is introduced in Section 4, wherein a summary of combinatorial characterizations of

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

optimality conditions for various choices of objective functions is also provided. Section 5 gives brief descriptions of Gauss maps and linear maps that will be exploited later in this paper. Section 6 proves the optimality conditions for fitting lines and planes to establish datums. Convexity of the objective functions is the subject of Section 7. Thus the major contributions of this paper are contained in Sections 6 and 7. Finally, Section 8 summarizes the results of the paper and offers some directions for future research.

2. Motivating Examples

Consider the problem of specification and verification involving datum planes. Figure 1(a) shows how a designer may graphically present the specification of position tolerancing of a cylindrical hole in a part with respect to a system of primary and secondary planar datums. Figure 1(b) illustrates how such a system of primary and secondary datum planes may be established on a manufactured instance of the part.

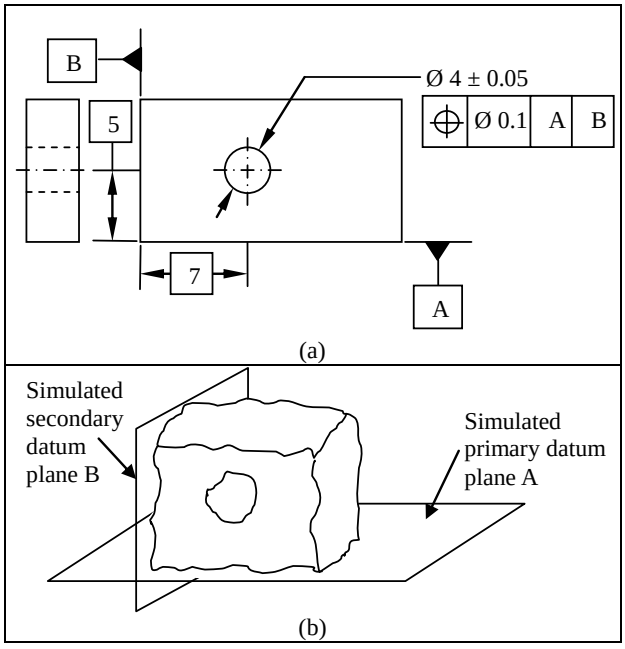


Figure 1. A simple example of (a) specification during design of a part, and (b) establishment of a system of primary and secondary datum planes on a manufactured instance of the part. The secondary datum plane B is required to be perpendicular to the primary datum plane A.

In the current era of digital manufacturing, a manufactured part, such as the one shown in Fig. 1(b), can be scanned by a CMS to generate a large number of discrete points, each with three-dimensional Cartesian coordinates. Those points corresponding to the datum feature A can then be processed to fit a primary datum plane A, and similarly for a secondary datum plane B. (For the sake of simplicity, a tertiary datum is not shown in Fig. 1.) The optimization problem for performing the plane fitting can be defined in several ways, depending on the choice of the objective function. In all cases, the datum planes are required to lie outside the material of the manufactured part while remaining as close to the part as possible (where 'as close as possible' is determined by the objective function).

It is also possible to specify datum planes for slabs and slots, as indicated in Fig. 2. Here two sets of points, each measured on each of the two features that correspond to the parallel plane features, are subjected to fitting by two parallel planes; the datum plane is the *median* plane of the two parallel planes that are thus fitted. In both cases involving slabs and slots, the fitted parallel planes are required to lie outside the material of the manufactured part while remaining as close to the part as possible. These median planes can also be used as secondary (or tertiary) datums; in that case additional constraints, such as the secondary datum being perpendicular to a primary datum, will be enforced.

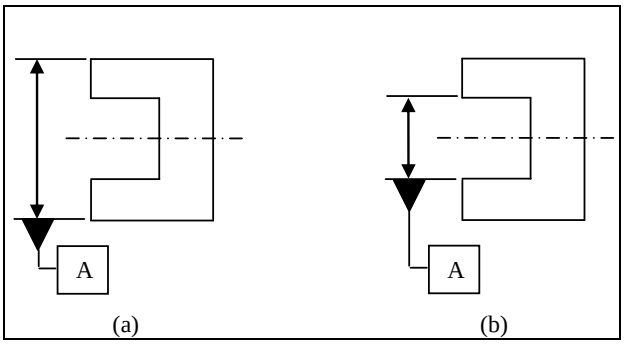


Figure 2. Specification of a datum plane for (a) a slab (external width) and (b) a slot (internal width). In each case, the datum is the median plane (indicated by dashes and dots) of two parallel planes that are fitted to two sets of points on two planar surface features (indicated by extension lines) on a manufactured instance of the part.

When a datum plane is a primary datum, the fitting is an optimization problem involving input points in space. This will also be referred to as a 3D (three-dimensional) problem in this paper. When a datum plane is a secondary datum, as datum B shown in Fig. 1, the fitting may be reduced (by projecting input points onto the primary datum plane) to an optimization problem involving points that lie on a (primary datum) plane; in this case, it is a line fitting problem for a set of input points in a plane. This will be referred to as a 2D (two-dimensional) problem in this paper. So in Fig. 1(b), the set of points measured on a manufactured part that correspond to the datum feature B are projected perpendicularly onto the datum plane A, to provide the input set of points for fitting a line.

A similar situation arises for establishing secondary datums involving slabs and slots. Here again, measured points on a manufactured part can be projected perpendicularly to a

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

primary datum plane, and parallel lines will be fitted to the two sets of input points in a plane. Therefore, the 2D problems of fitting lines and parallel lines to input points in a plane will be treated as equally important as 3D problems in the rest of the paper. In fact, several illustrations will deal with such line fittings in 2D while paying careful attention to the extendibility of the ideas to plane fittings in 3D.

3. Formulation of the Problem

Consider an arbitrary, continuous surface patch S in space and a plane P as shown in Fig. 3(a). Here $d(p,P)$ indicates the perpendicular distance of any point p in S to P. An infinitesimal area element around p is indicated as dA. Similar notations for an arbitrary, continuous curve C and a line L in a plane are shown in Fig. 3(b). Here $d(p,L)$ indicates the perpendicular distance of any point p in C to L. An infinitesimal line element around p is indicated as dl. It is assumed that the material of a manufactured part under consideration lies to one side of the surface patch S (and, similarly, to one side of the curve C). So when the plane P (or line L) is said to lie outside of the material, it should be apparent which side of S (or C) the plane P (or line L) should lie.

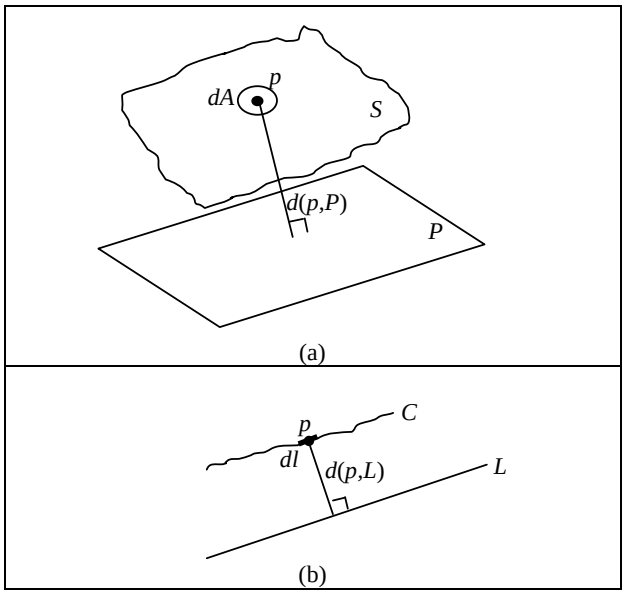


Figure 3. Notations for (a) fitting a plane, and (b) fitting a line.

A similar set of notations can be defined for parallel planes and parallel lines. For the sake of simplicity, only the case of parallel lines in a plane is illustrated in Fig. 4; the notations for the case of parallel planes in space can be inferred easily from this figure. Referring to Fig. 4, $d(p_1,L_1)$ denotes the perpendicular distance of any point p_1 in C_1 to line L_1, and $d(p_2,L_2)$ denotes the perpendicular distance of any point p_2 in C_2 to line L_2.

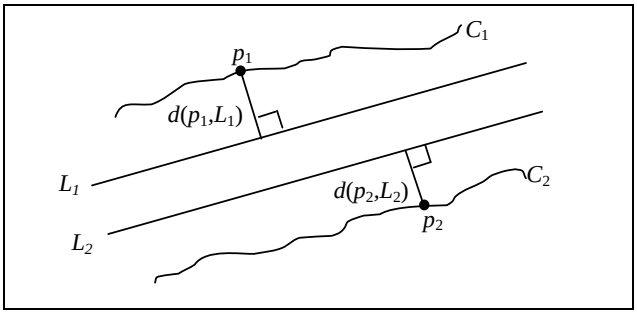


Figure 4. Notations for fitting parallel lines.

With these notations, and referring to Fig. 3(a), the constrained least-squares plane fitting problem can be posed as the following optimization problem:

$$\min_{P} \int_{S} d^2(p,P)\,dA \tag{1}$$

subject to P lying outside the material.

Similarly, referring to Fig. 3(b), the constrained least-squares line fitting problem can be posed as

$$\min_{L} \int_{C} d^2(p,L)\,dl \tag{2}$$

subject to L lying outside the material.

Posing the optimization problems for constrained least-squares fitting of parallel planes and parallel lines follows a similar pattern. For parallel planes P_1 and P_2, it is

$$\min_{P_1,P_2} \left[\int_{S_1} d^2(p_1,P_1)\,dA + \int_{S_2} d^2(p_2,P_2)\,dA \right] \tag{3}$$

subject to P_1 and P_2 being parallel, and both P_1 and P_2 lying outside the material.

For parallel lines L_1 and L_2, it is

$$\min_{L_1,L_2} \left[\int_{C_1} d^2(p_1,L_1)\,dl + \int_{C_2} d^2(p_2,L_2)\,dl \right] \tag{4}$$

subject to L_1 and L_2 being parallel, and both L_1 and L_2 lying outside the material.

The search space for planes and lines in the optimization problems posed in Eqs. (1) through (4) can be simplified considerably by the fact that the optimal planes and lines must contact at least one point of the input set of surface patches and

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

curves. For, if not, the plane (line) could be moved parallel to itself towards the input set while reducing the objective function of Eqs. (1) through (4), till it contacts at least one point of the input set. Such planes and lines have a special name. A plane (line) that contacts at least one point of an input set, which can be continuous or discrete, and keeps all the other points of the input set either on it or to one side of it, is called a *supporting* plane (line) of the input set [16]. Hence a simple result that gives a necessary condition for optimality follows:

Lemma 0: Solutions to the optimization problems posed in Eqs. (1) through (4) must be supporting planes or supporting lines to the input set of surface patches or curves, respectively.

Supporting lines and supporting planes generalize the concept of tangents to both discrete and continuous sets of points. The simple Lemma 0 also serves as the first combinatorial characterization of the optimality condition that reduces the search space. There is another subtle, but important, condition that can be associated with the supporting lines and planes. It is that the supporting line L (or supporting plane P) also keeps the immediate, infinitesimal material neighborhoods of the input curve (or surface patch) either on or to one side of L (or P). This removes any ambiguity as to which supporting lines and planes lie *outside* the material, and cuts down the search space of supporting lines and planes even further.

4. Convex Hull Filters and Comparison of Various Objective Functions

The optimization problems posed in Section 3 are quite appropriate for datum specification purposes during the design phase, when only perfect or imperfect continuous geometric entities are considered. When manufactured parts are measured using CMS for the purpose of verification, such lofty continuous objectives can seldom be realized. The output from CMS is usually a set of three-dimensional coordinates of discrete points, which may number in the millions. A practical and ingenious way to bridge the discrete world of CMS to the continuous world of surface patches and curves defined in Section 3 has been devised by the ISO and ASME standards committees using the convex hull filter. It is worth noting that the choice of convex hull for filtering is strongly motivated by Lemma 0.

Let $S = \{p_1, p_2, \dots, p_n\}$ be a set of input points from measurements made on a manufactured surface patch or a curve. For the sake of theoretical convenience, and without any loss of generality, let these input points lie 'more or less' on a horizontal plane or a horizontal line. Also let $CH(S)$ be the convex hull of S. Depending on the dimension, $CH(S)$ can be formed as either (in 3D) a convex polyhedron with triangular faces and hull edges that are line segments, or (in 2D) a convex polygon. If the input points are in general position, then $CH(S)$ will consist of hull vertices, hull edges that are line segments, and hull faces that are triangles. (If the points are not in general position, then the hull edges and faces can be partitioned into line segments and triangles, respectively.)

Consider the 2D problem first. If $CH(S)$ is a convex polygon in a plane, then every point of a hull edge (besides the end points) has one and only one supporting line – namely the one that contains that edge. There can be an infinite number of supporting lines through a hull vertex; the outward-pointing unit normals to these supporting lines form a wedge (a two-dimensional cone) with the hull vertex at the apex. However, not all the supporting lines will be candidates mentioned in Lemma 0. Only those supporting lines that lie outside the material will qualify for further consideration. This allows a partitioning of $CH(S)$ into an *outer* part of the convex hull (which is simply referred to in this paper as the 'outer convex hull') and an *inner* part of the convex hull ('inner convex hull'), with only the outer convex hull contributing to the search space of supporting lines referred to in Lemma 0.

A similar observation can be made if $CH(S)$ is a convex polyhedron in space. Every point in the interior of a triangular hull face has one and only one supporting plane – namely the one that contains that face. There can be an infinite number of supporting planes through a hull edge or hull vertex. In each case, the outward-pointing unit normals to these supporting planes form a wedge or a cone with the corresponding hull edge or hull vertex at the apex. Here again, using the material neighborhoods, a partitioning of $CH(S)$ into an outer part of the convex hull and an inner part of the convex hull can be effected, with only the outer convex hull contributing the search space of supporting planes referred to in Lemma 0.

The vertices of an outer convex hull of a given set of input set S of points form a subset of S. The rest of the points in S need not be considered further for the purpose of datum establishment. With this fact in mind, it is possible to compare the optimality conditions for various optimization criteria that have been studied in literature and practiced in industry. Table 1 gives a summary of the combinatorial characterizations of the optimality conditions for fitting lines and parallel lines in terms of the minimum number of contact with the outer convex hull(s). Table 2 provides a similar summary for planes and parallel planes. In the case of parallel lines and parallel planes, they are 'inscribed' for slots and 'circumscribed' for slabs.

In Tables 1 and 2, the indicated minimum contacts with the outer convex hulls occur when the input points are in general position. In degenerate cases, where the points are not in general position, more faces, edges, and vertices can contact the fitted planes. The utility of the results summarized in Tables 1 and 2 lies in the fact that these results may significantly reduce the search space to find solutions to the optimization problems of Eqs. (1) through (4). So it is worth discussing these optimization problems in some detail.

Various optimization problems shown in Tables 1 and 2 have a long and colorful history in engineering practice. The optimality conditions displayed in these tables have been exploited in commercial software to deliver practical solutions to industry. The optimization problems designated as CL_1P and CL_1L provide solutions that are in harmony with the 3-2-1 contacts for primary, secondary, and tertiary datum plane system, popularized in the engineering folklore [1]. Such

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

popularity is due to the apparent mechanical stability offered by the contacting planes to the manufactured parts. These plane and line fittings are also known as fitting under the L_1-norm.

Table 1. Comparison of optimality conditions for constrained least-squares fitting of lines and parallel lines to the outer convex hull(s).

	Designation	Optimization problem	Minimum contact with outer convex hull(s)	Select Ref.
Lines	CL_1L	Minimize integral of the distances to the supporting line.	edge	[17]
	SL_2L	Minimize integral of the square of the distances to a line without any constraint, then shift the fitted line to the outermost point(s).	vertex	[13,18]
	MZL	Minimize the maximum distance to the supporting line.	vertex	[19]
	CL_2L	Minimize integral of the square of the distances to the supporting line, per Eq. (2).	vertex or edge	[14,15, *]
Parallel lines	MILL	Maximize the distance between two parallel, inscribing lines.	vertex on each line.	[20]
	MCLL	Minimize the distance between two parallel, circumscribing lines.	vertex on one line and edge on the other line.	[19]
	SL_2LL	Minimize the integral of squares of distances to two parallel lines, and shift them to the outermost point(s).	vertex on each line.	[13]
	CL_2ILL	Minimize the sum of integrals of squares of distances to two parallel, inscribing lines, per Eq. (4).	(1) vertex on each line, or (2) vertex on one line and edge on the other line.	[*]
	CL_2CLL	Minimize the sum of integrals of squares of distances to two parallel, circumscribing lines, per Eq. (4).	(1) vertex on each line, or (2) vertex on one line and edge on the other line.	[*]

* indicates this paper.

The optimization problems designated as MZP and MZL have been the default definitions for datum planes and lines in ISO standards. These plane and line fittings are also known as fitting under the L_∞-norm [19]. SL_2P and SL_2L employ shifting a (total) least-squares fitting, also known as fitting under the L_2-norm, to the outermost point(s); these are well-known in research literature and lend themselves to elegant software implementations [13, 18]. Such least-squares fitting are also known for their numerical stability; that is, small changes in the input set of points result in only small changes in the fitted lines and planes. Their constrained counterparts, namely CL_2P and CL_2L, are relatively new and are the focus of recent research [14, 15] and this paper. These constrained least-squares fittings seem to combine the benefits of mechanical stability and numerical stability.

Table 2. Comparison of optimality conditions for constrained least-squares fitting of planes and parallel planes to the outer convex hull(s).

	Designation	Optimization problem	Minimum contact with outer convex hull(s)	Select Ref.
Planes	CL_1P	Minimize integral of the distances to the supporting plane.	face	[17]
	SL_2P	Minimize integral of the square of the distances to a plane without any constraint, then shift the fitted plane to the outermost point(s).	vertex	[13,18]
	MZP	Minimize the maximum distance to the supporting plane.	vertex or edge	[19]
	CL_2P	Minimize integral of the square of the distances to the supporting plane, per Eq. (1).	vertex, edge, or face	[14,15, *]
Parallel planes	MIPP	Maximize the distance between two parallel, inscribing planes.	vertex on each plane	[20]
	MCPP	Minimize the distance between two parallel, circumscribing planes.	(1) vertex on one plane and face on the other plane, or (2) edge on each plane	[19]
	SL_2PP	Minimize the integral of squares of distances to two parallel planes, and shift them to the outermost point(s).	vertex on each plane	[13]
	CL_2IPP	Minimize the sum of integrals of squares of distances to two parallel, inscribing planes, per Eq. (3).	(1) vertex on each plane, or (2) edge on each plane, or (3) vertex on one plane and edge on the other plane, or (4) vertex on one plane and face on the other plane.	[*]
	CL_2CPP	Minimize the sum of integrals of squares of distances to two parallel, circumscribing planes, per Eq. (3).	(1) vertex on each plane, or (2) edge on each plane, or (3) vertex on one plane and edge on the other plane, or (4) vertex on one plane and face on the other plane.	[*]

* indicates this paper.

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

A similar comparison can be made for fitting parallel planes and lines, as shown in Tables 1 and 2. MIPP and MILL have been the default fitting for slots (as shown in Fig. 2(b) for internal widths) in ASME and ISO standards. These optimization problems can be posed as quadratic programming problems [20]. MCPP and MCLL have also enjoyed the default status for fitting in ASME and ISO standards for slabs (as shown in Fig. 2(a) for external widths); these optimization problems can be posed as computations of widths of sets and are well studied in research literature [19]. SL_2PP and SL_2LL are the shifted versions of least-squares (under L_2-norm) fitting of parallel planes and parallel lines, which are otherwise unconstrained [13]. The problems designated as CL_2IPP, CL_2CPP, CL_2ILL, and CL_2CLL are relatively new, and are discussed in the rest of the paper along with CL_2P and CL_2L.

5. Gauss Maps and Linear Maps

Unconstrained (total) least-squares fitting of planes and lines to a set of discrete points or to an outer convex hull has been well explored in literature [13]. Results from these explorations can be extended to the constrained least-squares fitting of planes and lines, but these extensions require considerations of Gauss maps and linear maps, as described in Sections 5.1 and 5.2, respectively. These maps are combined to construct composite ellipsoids in Section 6 to provide the optimality conditions for constrained least-squares fitting.

5.1 Gauss Maps

Recall that supporting lines and supporting planes to the outer convex hulls form the feasible solutions to the optimization problems posed in Eqs. (1) through (4). There is a useful mapping, called the Gauss map, between these supporting planes (or lines) to points on a unit sphere (or circle) as shown in Fig. 5. Here, the outward-pointing unit normal of each supporting plane (or line) is mapped to a unique point on the unit sphere (or circle).

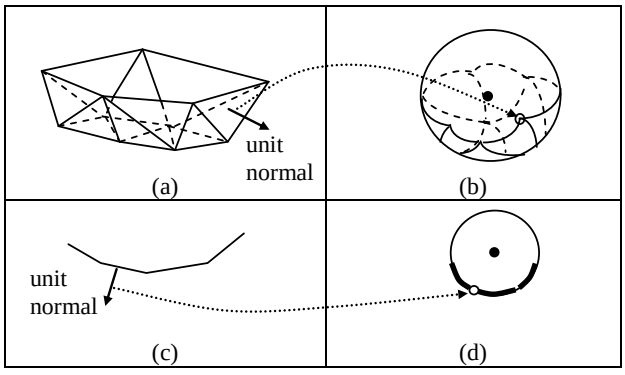


Figure 5. Gauss map obtained by mapping each unit normal of a supporting plane of an outer convex hull in space in (a) to a point on a unit sphere in (b). A similar Gauss map is obtained by mapping each unit normal of a supporting line of an outer convex hull in a plane in (c) to a point on a unit circle in (d).

For example, the outward-pointing unit normal v to a face of an outer convex hull in Fig. 5(a) is mapped to a point p on the unit sphere in Fig. 5(b), where the translated unit normal vector v starts at the center of the unit sphere and ends at the point p on the unit sphere. A similar mapping can be seen in Figs. 5(c) and 5(d), where an outward pointing unit normal v to an edge of an outer convex hull is mapped to a point p on a unit circle; the translated unit normal vector v starts at the center of the unit circle and ends at the point p on the unit circle.

The Gauss map partitions the unit sphere in Fig. 5(b) into faces, edges, and vertices that are (graph theoretic) dual to the vertices, edges, and faces, respectively, of the outer convex hull in Fig. 5(a). Similarly, the Gauss map partitions the unit circle in Fig. 5(d) into edges and vertices that are dual to the vertices and edges, respectively, of the outer convex hull in Fig. 5(c). Invoking the analogy of a global map on a sphere, the Gauss map in Figs. 5(b) and 5(d) can be viewed as covering the 'southern hemisphere' that includes the South Pole. If the outer convex hull in Figs. 5(a) and 5(c) were to be turned downside up, the Gauss map can be viewed as covering the 'northern hemisphere' that includes the North pole.

The incidence and adjacency relationships of the vertex-edge-face graph structure of the convex hull are preserved in the adjacency and incidence relationships of the face-edge-vertex graph structure in the Gauss map. The Gauss map is unique (injective, that is, one-to-one) for convex hulls, and is a powerful conceptual and computation tool to discuss the optimization problem of constrained least-squares fitting.

The topic of Gauss map is general, and is well developed in the mathematical literature. The Gauss map as visualized on a unit sphere has an appealing geographical metaphor in the form of a globe partitioned into regions. Additional notions such as the equator, and the North and South Poles on the unit sphere help the conceptual development of solutions to the optimization problem later in this paper.

5.2 Linear Maps

A linear map M is a linear transformation of vectors from $\mathbb{R}^n$ to $\mathbb{R}^m$. It can be represented by a matrix $M \in \mathbb{R}^{m\times n}$. Let $m > n$ for the purpose of this paper. The singular value decomposition (SVD) of M is given by

$$M = U\Sigma V^T, \tag{5}$$

where $U \in \mathbb{R}^{m\times m}$ is a left singular matrix, $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix of singular values, and $V \in \mathbb{R}^{n\times n}$ is a right singular matrix [21-23]. Both U and V are ortho-normal matrices. By convention, the n singular values $\{\sigma_i, i = 1, \dots, n\}$ occupy the diagonal of the top $n \times n$ block of Σ and the remaining cells of Σ are filled with zeros. Also, the right m-n columns of U are filled with arbitrary ortho-normal columns to maintain the relation $U^T U = I$.

The particular relationship between a left singular vector u_j and a right singular vector v_j can be obtained from Eq. (5) as

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

$$Mv_j = \sigma_j u_j. \quad (6)$$

The SVD of M is closely related to the eigenvalue problem of $M^{\mathrm{T}}M$. The eigenvalues of $M^{\mathrm{T}}M$ are the squares of the singular values of M, and the eigenvectors of $M^{\mathrm{T}}M$ are the right singular vectors of M. So it follows that

$$(M^{\mathrm{T}}M)v_j = \sigma_j^2 v_j. \quad (7)$$

A geometrical interpretation of the SVD in Eq. (5) is given by the following result [23].

Theorem 1: A linear map $M \in \mathbb{R}^{m\times n}$ transforms a unit sphere in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^m$. The principal axes of the ellipsoid are aligned with the left singular vectors in U and the intercepts (semi-axes) of the ellipsoid with the principal axes are given by the singular values in Σ.

Figure 6 illustrates the geometric interpretation given by Theorem 1 when M is a 2×2 matrix. Under this linear map M a unit sphere (in this case, a unit circle) is mapped to an ellipsoid (in this case, an ellipse). The singular values and singular vectors of M are related by Eq. (6) as $Mv_1 = \sigma_1 u_1$ and $Mv_2 = \sigma_2 u_2$. In general, the right singular vectors are the preimages of the principal semi-axes of the ellipsoid realized in the left singular vector space.

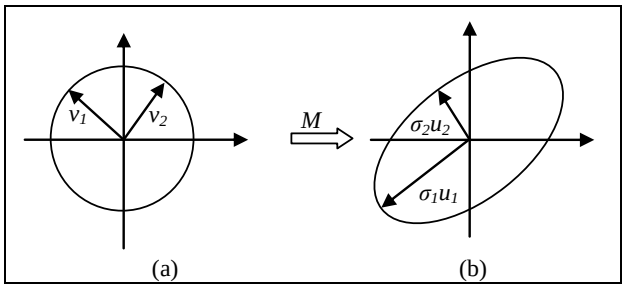


Figure 6. Geometrical illustration of the linear mapping M of (a) a unit circle to (b) an ellipse using the singular value decomposition of M.

Although the ellipsoidal nature of the linear map in the left singular vector space, as illustrated in Fig. 6(b), is very appealing, the right singular vector space is more useful for further discussion. As Eq. (7) suggests, the eigenvectors of $M^{\mathrm{T}}M$ live in the right singular vector space of M. A combination of Theorem 1 and Eq. (7) leads to the following important result for optimization.

Theorem 2: The function $v^{\mathrm{T}}M^{\mathrm{T}}Mv$ of unit vectors v on the unit sphere in $\mathbb{R}^n$ reaches a minimum value of σ_j^2 at $v = v_j$, where σ_j is the smallest singular value of M and v_j is the corresponding right singular vector.

An interesting and useful geometrical interpretation of Theorems 1 and 2 is provided by the following result in the right singular vector space.

Theorem 3: The spherical plot of the function $\sqrt{v^{\mathrm{T}}M^{\mathrm{T}}Mv}$ of unit vectors v defined on the unit sphere in $\mathbb{R}^n$ is an ellipsoid in $\mathbb{R}^n$, with the principal axes of the ellipsoid aligned with the right singular vectors of M and the semi-principal axes assuming the singular values of M.

The ellipsoid (and its specialization to an ellipse in 2D) of Theorem 3 will play a very important role in the theoretical developments of this paper.

All the results presented on linear maps in this section are general, and are well established in literature [21-23]. The main advantage of Theorem 3 is that all geometrical reasoning can be based on ellipses in 2D and ellipsoids in 3D for the constrained least-squares fitting problem addressed in the following section.

6. Optimality Conditions

The minimization achieved in Theorem 2 is made possible by analyzing a single ellipsoid described in Theorem 3. For the constrained minimization problems addressed in this paper, no single ellipsoid is sufficient. It is shown in this section that a composite ellipsoid, which is a continuous surface consisting of several ellipsoidal patches, is needed. The Gauss map and the linear map described in Section 5.1 and 5.2 will be used in constructing such a composite ellipsoid, based on some real matrices.

The case of fitting single lines and planes will be discussed in Section 6.1. The results will be extended in Section 6.2 to the case of fitting of pairs of lines and pairs of planes, where the pairs may intersect or be parallel. From now on, it should be noted that the notion of 'objective function' will implicitly include the constraints as well. This is because Lemma 0 enables automatic incorporation of the constraints by restricting the planes and lines in Eqs. (1) through (4) to be supporting planes and supporting lines, respectively.

6.1 Single Lines and Planes

The type of real matrix M dealt with in the rest of this paper will have two important features. First, it has either three columns (n = 3) when the fitting is done in three-dimensional space, or two columns (n = 2) when the fitting takes place in a plane. The second feature is that the origin of the coordinate system involved in defining the matrix M will be shifted, by pure translation, to a convenient point p for mathematical and computational purposes. To make this clear, the notation M_p will be used to indicate the point p that is used as the origin to construct that particular matrix M_p.

With these preliminaries, the objective functions defined in Eqs. (1) and (2) can be obtained by integration over the outer convex hulls of Figs. 5(a) and 5(c) treated as a surface and as a curve, respectively. For this, consider the origin of the coordinate system of the CMS to be an arbitrary point O in 3D and a plane P (or line L in 2D case) through O that has a unit

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

normal v. Then the integral of the square of the distance of a point on the outer convex hull from the plane P (or line L) can be expressed as the square of the second-norm of a linear map expanded as

$$\|M_O v\|_2^2 = (M_O v)^{\mathrm{T}}(M_O v) = v^{\mathrm{T}} M_O^{\mathrm{T}} M_O v, \tag{8}$$

where M_O is a matrix that will be described in some detail in the following Eqs. (9) and (10). Note the striking similarity between the objective function expressed in Eq. (8) and the function whose spherical plot is described in Theorem 3. Mathematically speaking, minimizing an objective function (as in Eq. (8) and Theorem 2) is the same as minimizing its square-root (as in Theorem 3).

In three-dimensional cases, it has been shown [14] that an exact integration of the expression in Eq. (1) over an outer convex hull, such as the one in Fig. 5(a), can be carried out using the Simpson's rule (over triangles in 3D) so that

$$M_O = \sqrt{\frac{1}{12}} \begin{bmatrix} \sqrt{A_1}x_{1A} & \sqrt{A_1}y_{1A} & \sqrt{A_1}z_{1A} \\ \sqrt{A_1}x_{1B} & \sqrt{A_1}y_{1B} & \sqrt{A_1}z_{1B} \\ \sqrt{A_1}x_{1C} & \sqrt{A_1}y_{1C} & \sqrt{A_1}z_{1C} \\ 3\sqrt{A_1}\bar{x}_1 & 3\sqrt{A_1}\bar{y}_1 & 3\sqrt{A_1}\bar{z}_1 \\ \vdots & \vdots & \vdots \\ \sqrt{A_N}x_{NA} & \sqrt{A_N}y_{NA} & \sqrt{A_N}z_{NA} \\ \sqrt{A_N}x_{NB} & \sqrt{A_N}y_{NB} & \sqrt{A_N}z_{NB} \\ \sqrt{A_N}x_{NC} & \sqrt{A_N}y_{NC} & \sqrt{A_N}z_{NC} \\ 3\sqrt{A_N}\bar{x}_N & 3\sqrt{A_N}\bar{y}_N & 3\sqrt{A_N}\bar{z}_N \end{bmatrix}, \tag{9}$$

where the outer convex hull is comprised of N triangles T_i, each having area A_i and vertices (x_{iA}, y_{iA}, z_{iA}), (x_{iB}, y_{iB}, z_{iB}), and (x_{iC}, y_{iC}, z_{iC}), their average being ($\bar{x}_i$, $\bar{y}_i$, $\bar{z}_i$).

Similarly, in two-dimensional cases it has been shown [14] that the application of Simpson's rule results in an exact integration of the expression in Eq. (2) over an outer convex hull, such as the one in Fig. 5(c), yielding the $3N \times 3$ matrix

$$M_O = \sqrt{\frac{1}{6}} \begin{bmatrix} \sqrt{L_1}(x_1) & \sqrt{L_1}(y_1) \\ 2\sqrt{L_1}\left(\frac{x_1+x_2}{2}\right) & 2\sqrt{L_1}\left(\frac{y_1+y_2}{2}\right) \\ \sqrt{L_1}(x_2) & \sqrt{L_1}(y_2) \\ \vdots & \vdots \\ \sqrt{L_N}(x_N) & \sqrt{L_N}(y_N) \\ 2\sqrt{L_N}\left(\frac{x_N+x_{N+1}}{2}\right) & 2\sqrt{L_N}\left(\frac{y_N+y_{N+1}}{2}\right) \\ \sqrt{L_N}(x_{N+1}) & \sqrt{L_N}(y_{N+1}) \end{bmatrix}, \tag{10}$$

where the outer convex hull is comprised of N line segments each having length L_i, and $N+1$ vertices each with coordinates (x_i, y_i).

The matrices in Eqs. (9) and (10) can be used to construct composite ellipsoids in 3D and composite ellipses in 2D, respectively. For simplicity of exposition, the construction of a composite ellipse is described first in some detail. Figure 7 illustrates an outer convex hull, its Gauss map, and a composite ellipse that consists of several elliptic arcs. Reference to Figs. 5(c) and 5(d) can be made for comparison.

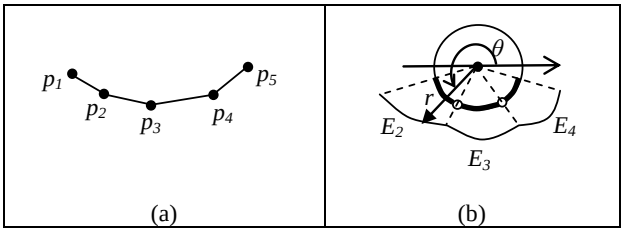


Figure 7. Illustration of (a) an outer convex hull, and (b) a composite ellipse using polar plots over a Gauss map.

The vertices of the outer convex hull in Fig. 7(a) are indicated as $\{p_1, p_2, \ldots, p_5\}$. The Gauss map of the outer convex hull is shown in Fig. 7(b) on a unit sphere (in this case, a unit circle), where the three thick circular arcs correspond to the cones of normals at vertices $p_2, p_3,$ and p_4. Now consider the vertex p_2 and the matrix M_{p_2} defined in Eq. (10). Here, note that the origin for the coordinate system is translated to the vertex p_2 to compute the coordinates involved in M_{p_2}. Then Theorem 3 defines an ellipse, and the elliptic arc E_2 in Fig. 7(b) is the restriction of this ellipse to the Gauss map corresponding to the cone of outer normals at the vertex p_2. Note that the elliptic arc E_2 is drawn as a polar plot of $r = \sqrt{v^{\mathrm{T}} M_{p_2}^{\mathrm{T}} M_{p_2} v}$ as a function of the angle θ between the unit vector v and the horizontal axis, as shown in Fig. 7(b). The other elliptic arcs E_3 and E_4 in Fig. 7(b) are constructed similarly using polar plots to obtain the composite ellipse.

Once the composite ellipse is constructed, the optimization problem reduces to finding the unit vector v (or, equivalently, a point on the unit circle) in Fig. 7(b) that yields the smallest radius in the polar plot of the composite ellipse, and a corresponding point p on the outer convex hull in Fig. 7(a) for which v is the outer normal. Then, the constrained least-squares fitting line is uniquely found as the one passing through p and having v as its normal.

This leads to only two possibilities for the optimality condition in 2D, as described in the following case analyses.

Case 1: The minimum radius of the polar plot is realized at an interior point of an elliptic arc. In this case, the constrained least-squares line contacts only the vertex of the outer convex hull responsible for that elliptic arc.

Case 2: The minimum radius of the polar plot occurs at the intersection of two adjacent elliptic arcs. In this case, the constrained least-squares line contacts the edge of the outer

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

convex hull that joints the two vertices that correspond to the two adjacent elliptic arcs.

Thus, the optimality condition for the 2D problem is given by the following theorem.

Theorem 4a: The constrained least-squares line for a set of input points in a plane can contact one or two points of the input set. These contact points correspond to a vertex or an edge of the outer convex hull.

A similar characterization for fitting a plane to points in space can be obtained by constructing a composite ellipsoid as an extension of Fig. 7. Starting with an outer convex hull, such as the one in Fig. 5(a), the composite ellipsoid is constructed as a spherical plot using the unit sphere, the Gauss map, and Theorem 3. The composite ellipsoid consists of a continuous collection of ellipsoidal patches $\{E_i\}$, each E_i corresponding to a vertex p_i on the outer convex hull in 3D. Here E_i arises from the ellipsoid of Theorem 3 constructed with the matrix M_{p_i} from Eq. (9). The ellipsoid itself can be constructed as a spherical plot of the radius $r = \sqrt{v^{\mathrm{T}} M_{p_i}^{\mathrm{T}} M_{p_i} v}$ as a function of two angles (say, the latitude θ and longitude φ) associated with the unit vector v in the unit sphere. E_i is then the restriction of this ellipsoid to the Gauss map corresponding to the cone of outer normals at the vertex p_i.

In seeking the minimum radius of the spherical plot that represents the composite ellipsoid, there are only three possibilities, as described in the following case analyses.

Case 1: The minimum radius of the spherical plot is realized at an interior point of an ellipsoidal patch. In this case, the constrained least-squares plane contacts only the vertex of the outer convex hull responsible for that ellipsoidal patch.

Case 2: The minimum radius of the spherical plot occurs at the intersection of two adjacent ellipsoidal patches. In this case, the constrained least-squares plane contacts the edge of the outer convex that joints the two vertices that correspond to the two adjacent ellipsoidal patches.

Case 3: The minimum radius of the spherical plot occurs at the intersection of three adjacent ellipsoidal patches. In this case, the constrained least-squares line contacts the triangular face of the outer convex that is defined by the three vertices that correspond to the three adjacent ellipsoidal patches.

Thus, the optimality condition for the 3D problem is given by the following theorem.

Theorem 4b: The constrained least-squares plane for a set of input points in space can contact one, two, or three points of the input set. These contact points correspond to a vertex, an edge, or a face of the outer convex hull.

Theorems 4(a) and 4(b) form the foundations for some of the optimality conditions enumerated in Tables 1 and 2. In particular, comparisons can be made between CL_2L and all the other line fittings (CL_1L, SL_2L, and MZL) in Table 1. Similarly, comparisons can be made between CL_2P and all the other plane fittings (CL_1P, SL_2P, and MZP) in Table 2. The advantage of the constrained least-squares fitting over the others is that it combines the benefits of mechanical stability (responsible for the popularity of CL_1L and CL_1P) and numerical stability (responsible for the popularity of SL_2L and SL_2P). This is perhaps the strongest reason for its attraction to the ASME and ISO standards community.

The process of establishing the optimality conditions of Theorems 4(a) and 4(b) has also revealed some elegant algorithmic ideas that can be used to find the optimum solution. Construction of outer convex hulls, their Gauss maps, the linear maps of Eqs. (9) and (10), and the composite ellipsoids/ellipses can all be part of an algorithm to find the constrained least-squares fitting solution.

6.2 Pairs of Lines and Pairs of Planes

The problem of fitting more than one single plane or one single line occurs while establishing datums for slabs and slots. Also, especially in ISO standards [3, 4], datums can be established for wedges and angular slots without regard to the included angle. To address these needs, Section 6.2.1 tackles the problem of fitting intersecting lines and intersecting planes. Section 6.2.2 takes up the problem of fitting parallel lines and parallel planes.

6.2.1 Intersecting Lines and Intersecting Planes

The constrained least-squares fitting of lines and planes can be used directly for establishing datum systems for wedges and angular slots. As shown in Fig. 8(a) for 2D cases, each curve of a solid wedge can be independently subjected to a constrained least-squares fitting of a line. The datum is then a system consisting of the median line (i.e., angle bisector) of the intersecting lines and a point (e.g., at the intersection of fitted lines) on the median line. A similar approach is illustrated in Fig. 8(b) for an angular slot. Note that such a datum system enables the establishment of a complete two-dimensional reference frame.

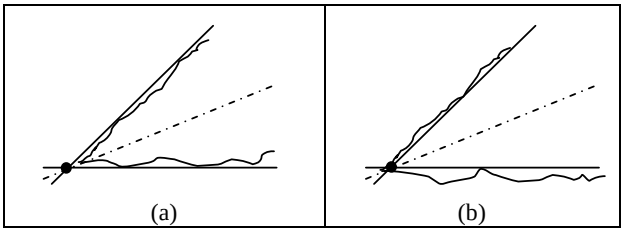


Figure 8. Illustration of datum for (a) a wedge, and (b) an angular slot. The datum is a system consisting of the median line indicated with dashes and dots, and a point indicated with a filled circle on the median line.

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

Extending these ideas to 3D cases of wedges and angular slots, a constrained least-squares plane can be fitted on each surface independently, resulting in intersecting planes. The datum is then a system consisting of the median plane (i.e., angle bisector) of the intersecting planes and a line (e.g., at the intersection of the fitted planes) on the median plane. Such a datum system will correspond to a prismatic class of datums [1, 4].

6.2.2 Parallel Lines and Parallel Planes

The constrained least-squares fitting of parallel lines and parallel planes can be treated in a manner similar to that of constrained least-squares fitting of single line and single plane. Starting with the optimization problems defined in Eqs. (3) and (4), the notions of outer convex hulls, Gauss maps, and linear maps can still be applied with the only additional constraint that the lines and planes be parallel. In addition, the Gauss map will have two connected components, one in the northern hemisphere and the other in the southern hemisphere of the unit sphere; also, the composite ellipsoid will have two connected components, one in each hemisphere. These two components correspond to the surfaces S_1 and S_2 in Eqn. (3), or to the curves C_1 and C_2 in Eqn. (4).

For simplicity of exposition, the 2D problem is considered first in some detail. Figure 9 illustrates a simple 2D example involving an upper convex hull and a lower convex hull as in Fig. 9(a). Both are outer convex hulls for the respective set of input points. The unit circle and the composite ellipse are shown in Fig. 9(b). The composite ellipse consists of two connected components of elliptic arcs.

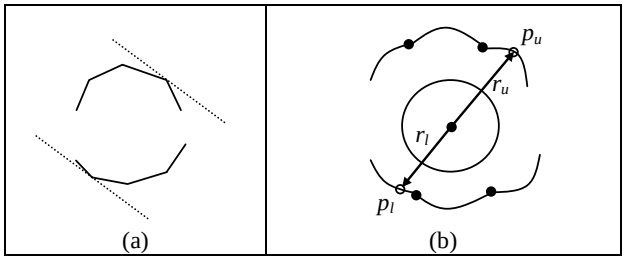


Figure 9. (a) An example of upper and lower convex hulls, and (b) the unit circle and the corresponding composite ellipse consisting of two connected components.

The two parallel supporting lines shown dotted in Fig. 9(a) define a normal direction along which two antipodal (that is, diametrically opposite) points p_u and p_l can be located on the composite ellipse as in Fig. 9(b). The corresponding radii of the composite ellipse are designated as r_u and r_l. The objective function of Eq. (4) is then given by $r_u^2 + r_l^2$. This leads to only two possibilities for the optimality condition in 2D, as described in the following case analyses.

Case 1: The minimum of $r_u^2 + r_l^2$ is realized when both the antipodal points p_u and p_l are at interiors of elliptic arcs. In this case, the constrained least-squares parallel lines contact a vertex from each of the outer convex hulls.

Case 2: The minimum of $r_u^2 + r_l^2$ is realized when one antipodal point is at an interior of an elliptic arc and the other antipodal point is at the intersection of two adjacent elliptic arcs. In this case, the constrained least-squares parallel lines contact a vertex on one outer convex hull and an edge on the other outer convex hull.

The possibility of realizing a minimum when both the antipodal points p_u and p_l are at the intersections of elliptic arcs is considered a special case, where the input points are in a degenerate position. Therefore, the edge-edge contact is not considered for the optimality conditions when the input points are in general position. Thus, the optimality condition for the 2D problem is given by the following theorem.

Theorem 5a: The constrained least-squares fitting of parallel lines to two sets of input points in a plane can have one contact point on each parallel line, or one contact point on one parallel line and two contact points on the other parallel line.

These contact points on the two outer convex hulls correspond to a vertex on each parallel line, or a vertex on one parallel line and an edge on the other parallel line. Table 1 summarizes these optimality conditions for CL_2ILL and CL_2CLL.

A similar set of case analyses can be conducted for the 3D problem of fitting parallel planes. Generalizing Fig. 9 to 3D results in a composite ellipsoid in 3D, with two connected components of ellipsoidal patches. Two parallel supporting planes, generalizing Fig. 9(a), will yield a common normal direction that defines two antipodal points p_u and p_l on the composite ellipsoid with two radii r_u and r_l. The objective function of Eq. (3) is then given by $r_u^2 + r_l^2$. This leads to only four possibilities for the optimality condition in 3D, as described in the following case analyses.

Case 1: The minimum of $r_u^2 + r_l^2$ is realized when both the antipodal points p_u and p_l are at interiors of ellipsoidal patches. In this case, the constrained least-squares parallel planes contact a vertex from each of the outer convex hulls.

Case 2: The minimum of $r_u^2 + r_l^2$ is realized when both the antipodal points are at interiors of elliptic arcs (each being the intersection of two adjacent ellipsoidal patches). In this case, the constrained least-squares parallel planes contact an edge from each of the outer convex hulls. Note that these two edges will form two skew lines in space.

Case 3: The minimum of $r_u^2 + r_l^2$ is realized when one antipodal point is at the interior of an ellipsoidal patch and the other antipodal point is at the interior of an elliptic arc (being

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

the intersection of two adjacent ellipsoidal patches). In this case, the constrained least-squares parallel planes contact a vertex on one outer convex hull and an edge on the other outer convex hull.

Case 4: The minimum of $r_u^2 + r_l^2$ is realized when one antipodal point is at the interior of an ellipsoidal patch and the other antipodal point at an ellipsoidal vertex (being the intersection of three adjacent ellipsoidal patches). In this case, the constrained least-squares parallel planes contact a vertex on one outer convex hull and a face on the other outer convex hull.

The possibility of realizing a minimum when both the antipodal points p_u and p_l are at the (ellipsoidal) vertices of the composite ellipsoid is considered a special case, where the input points are in a degenerate position. Therefore, the face-face contact is not considered for the optimality conditions when the input points are in general position. A similar consideration also rules out the edge-face contact. Thus, the optimality condition for the 3D problem is given by the following theorem.

Theorem 5b: The constrained least-squares fitting of parallel planes to two sets of input points in space can have one contact point on each parallel plane, or two contact points on each parallel plane, or one contact points on one parallel plane and two contact points on the other parallel plane, or one contact point on one parallel plane and three contact points on the other parallel plane.

These contact points on the two outer convex hulls correspond to a vertex on each parallel plane, or an edge on each parallel plane, or a vertex on one parallel plane and an edge on the other parallel plane, or a vertex on one parallel plane and a face on the other parallel plane. Table 2 summarizes these optimality conditions for CL_2IPP and CL_2CPP.

7. Convexity of Objective Functions

A blind application of the ideas outlined in Section 6 to find a global solution to the constrained least-squares fitting may lead to an exhaustive search over all the ellipsoidal patches of the composite ellipsoid. This can be avoided if the objective function is convex. Then only a local minimum is needed because it will also serve as the global minimum under convexity. This section shows that the objective functions are indeed convex even under the constraints of Eqs. (1) through (4). Convexity of the objective functions is indeed a critically important property for practical and efficient computation, especially in light of the numerous combinatorial conditions given by Theorems 4a, 4b, 5a and 5b, all of which would have be checked in an exhaustive search for a global minimum in the absence of convexity.

The notions of composite ellipse and composite ellipsoid are central to further discussion. As noted earlier, there is an important one-to-one correspondence among the following three entities: (1) an outward pointing normal *v* of a supporting plane (or line), (2) a point on the Gauss map etched on a unit sphere (or unit circle), and (3) a point on a composite ellipsoid (or ellipse). Recall that the faces, edges, and vertices of a Gauss map are dual to the vertices, edges, and faces of an outer convex hull. So the ellipsoidal patches (or elliptic arcs), ellipsoidal edges, and ellipsoidal vertices (or vertices between elliptic arcs) are dual to the vertices, edges, and faces of an outer convex hull.

While discussing the behavior of the objective function, frequent references will be made to points in the interior of an ellipsoidal patch and to points across (that is, at the intersection of) ellipsoidal patches to evoke a geometrical perception of the optimization problem. At those moments, imagining the correspondence of these points to the normals of supporting planes, or to points on Gauss map etched on a unit sphere, will greatly aid the comprehension of the theoretical arguments. In addition, references to the South Pole, southern hemisphere, and equatorial region will be made on the Gauss map on a unit sphere. Also, in much of the discussion, the relevant domain of interest for optimization will be a (fairly large) portion of the 'southern hemisphere' near the South Pole because that is where the solution of interest lies.

Armed with these observations, the objective functions will be shown to be convex by proving that they are convex in every neighborhood within the domain of interest. Section 7.1 considers the neighborhoods in the interior of elliptic arcs and ellipsoidal patches. Section 7.2 examines the convexity along the edges of intersection of the ellipsoidal patches. Section 7.3 then examines the neighborhoods across the elliptic arcs and ellipsoidal patches. Section 7.4 provides the convexity argument for fitting pairs of lines and pairs of planes.

7.1 Convexity in the Interior of Patches

Consider first the 2D problem of constrained least-squares fitting of lines in a plane. The composite ellipse shown in Fig. 7(b) can then be used as an example. Figure 10(a) reproduces Fig. 7(b) for the polar plot of the composite ellipse. In Fig. 10(b) a Cartesian plot of the integral of the squares of the distances (which is r^2) is presented as a function of the single degree of freedom θ. The goal is to establish that this objective function in Fig. 10(b) is convex over the entire domain of interest.

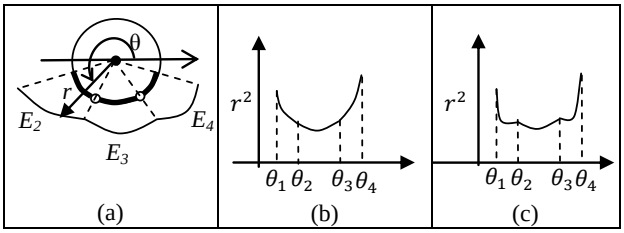


Figure 10. Illustrations of (a) a polar plot, (b) a Cartesian plot of the objective function as being convex, and (c) a non-convex objective function.

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

The elliptic arc E_2 of Fig. 10(a) is mapped in the interval (θ_1, θ_2) in Fig. 10(b) after squaring the radius function. Similar mappings of E_3 and E_4 can be seen in Fig. 10(b) in the subsequent intervals. In the interior of each of these intervals, such as (θ_1, θ_2), the radius r as a function of the angle θ can be seen to be convex because of the property of ellipse in the neighborhood of its minor axis (that is, near the South Pole of the unit circle in Fig. 10(a)). Since the square of a positive convex function is convex, the objective function r^2 is also convex in the interior of the interval (θ_1, θ_2). This result holds good in every one of the subsequent intervals shown in Fig. 10(b), thus establishing the following fact.

Lemma 1a: The objective function is convex at every neighborhood within each of the elliptic arcs in the composite ellipse over the domain of interest for the constrained least-squares fitting of lines in a plane.

The argument for the convexity of the objective function in the interior of the elliptic arcs can be extended to the convexity of the objective function in the interior of ellipsoidal patches in the 3D cases. Here again the radius r of an ellipsoid plotted as a Cartesian function of two degrees of freedom (say, the angles of longitude θ and latitude φ) is convex near the minor axis of the ellipsoid (that is, near the South Pole of the unit sphere). Therefore, the objective function r^2 must be convex in the interior of the patches in the θ-φ domain that correspond to the interior of the ellipsoidal patches, leading to the following fact.

Lemma 1b: The objective function is convex at every neighborhood within each of the ellipsoidal patches in the composite ellipsoid over the domain of interest for the constrained least-squares fitting of planes in space.

7.2 Convexity Along Edges of Patches

A composite ellipsoid will have elliptic arcs, with each arc as the intersection of two adjacent ellipsoidal patches. These intersection curves are pieces of planar curves (ellipses) and the radius function $r(\theta, \varphi)$ is convex in the southern hemisphere near the South Pole. So r^2 must also be convex in the domain of interest, leading to the following result.

Lemma 1c: The objective function is convex along the interior of every edge (that is, excluding its end points) of intersection of the ellipsoidal patches in the composite ellipsoid over the domain of interest for the constrained least-squares fitting of planes in space.

The objective function is convex even at the end points (that is, the vertices) of the edges of the ellipsoidal patches, as shown in the next section.

7.3 Convexity Across Patches

Again, consider the 2D problem first. Having established that the objective function is convex in the interiors of the elliptic arcs, it is now necessary to show that situations such as Fig. 10(c) do not occur *across* the elliptic arcs. For this, an interesting alternative expression of the objective function is helpful. This involves the consideration of the centroid g of the outer convex hull. Following the arguments presented in [14, 15] it can be shown that in 2D cases

$$v^{\mathrm{T}} M_p^{\mathrm{T}} M_p v = v^{\mathrm{T}} M_g^{\mathrm{T}} M_g v + L(|v^{\mathrm{T}}(p-g)|)^2, \qquad (11)$$

where L is the total length of the outer convex hull. Figure 11 provides an illustration to give a geometrical interpretation of Eq. (11) in 2D.

Consider the outer convex hull in Fig. 11 that is a reproduction of the one in Fig. 5(c). In Fig. 11 the centroid of the outer convex hull is indicated as g. Now consider a supporting line l_p for the outer convex hull through a hull vertex p and a unit normal vector v, and a line l_g that is parallel to l_p and passing through g. Let h be the distance between these two parallel lines l_p and l_g. Then, according to Eq. (11), the integral of the square of the distances of the convex hull from l_p is equal to the integral of the square of the distances of the convex hull from l_g plus the length of the convex hull times the square of the distance h between these two parallel lines. This property is sometimes referred to as the 'parallel axis theorem.' The point p can be any point in the plane and does not have to be a convex hull vertex for Eq. (11) to be valid.

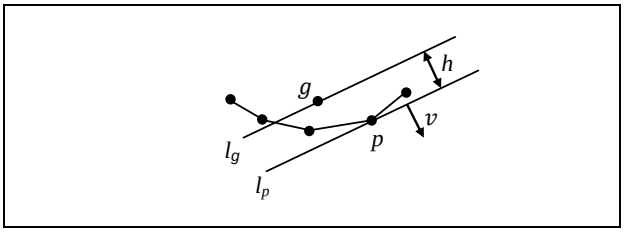


Figure 11. Illustration for the interpretation of Eq. (11).

Since the sum of convex functions is convex, the convexity of the objective function in Eq. (11) across the elliptic arcs is proved if each of the two terms on the right side of Eq. (11) is proved to be a convex function. Consider the first term $v^{\mathrm{T}} M_g^{\mathrm{T}} M_g v$ in the right side of Eq. (11). According to Theorem 3, the polar plot of its square-root is an ellipse, and its Cartesian plot with respect angle θ is convex and positive in the neighborhood of the minor axis (that is, near the South Pole). So its square is also convex even across the elliptic arcs of Fig. 10(a).

It then only remains to prove that the second term $L(|v^T(p-g)|)^2$ in the right side Eq. (11), which is the same as Lh^2, is convex across the elliptic arcs. It is possible to show that h^2 is a convex *across* the elliptic arcs (and thus Lh^2 is a convex function of the angle θ) using case analyses as illustrated in Fig. 12. In each of the three figures in the top row of Fig. 12, a continuous family of supporting lines passes through the vertex p_i and similarly another continuous family of supporting lines

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

passes through the vertex p_j. The transition from one elliptic arc to another occurs at the instant when the pivot for the family of supporting lines switches from p_i to p_j. That moment of transition is captured in Fig. 12 using three case analyses.

In particular, the case analyses of Fig. 12 involve the behavior of the perpendicular distance h from the centroid g to the supporting lines as a function of the rotation angle θ. Since the search for the minimum takes place in the (fairly large) vicinity of the South Pole, it can be first observed from the top row of Fig. 12 that h varies as $\cos\theta$ in the domain of interest. Therefore, the h^2-θ plots in the bottom row of Fig. 12 consist of pieces of cosine-square functions.

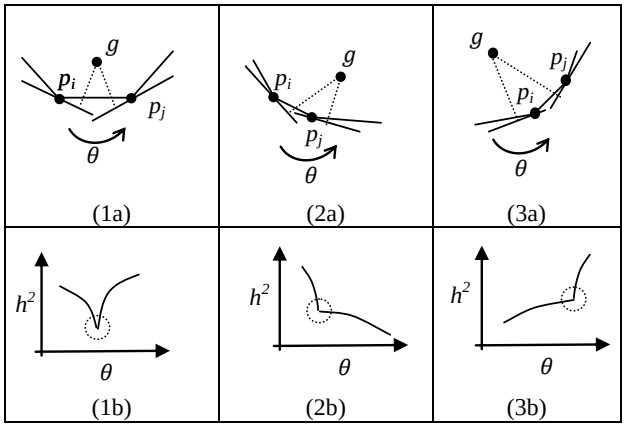


Figure 12. Illustration of three cases of transition and the related objective functions. The perpendiculars h from the centroid g to the supporting lines are shown dotted in the top row. The transition neighborhoods are shown within dotted circles in the h^2-θ plots in the bottom row.

Even though the pieces of cosine-square functions in the bottom row of Fig. 12 are themselves not convex, they behave like convex functions in the transition neighborhoods indicated within dotted circles again in the bottom row of Fig. 12. The concept of local convexity of a function is illustrated in Fig. 13. For a function *f(x)* to be convex over an arbitrary (including vanishingly small) interval (x_1, x_2) in Fig. 13, it is sufficient to show that the line-segment connecting any two points, such as *a* and *b*, on the graph of the function within that interval lies entirely above the graph of that function.

Figure 12 illustrates all the three cases of possible transitions. The first case illustrated in Figs. 12(1a) and 12(1b) shows how the perpendicular distance h from the centroid g to the supporting lines changes with respect to the rotation angle θ. As shown in the h^2-θ plot in Fig. 12(1b), the transition reaches a local minimum and is locally convex. The second case illustrated in Figs. 12(2a) and 12(2b) also shows that the transition in the h^2-θ plot is kept locally convex. The third case shown in Figs. 12(3a) and 12(3b) is similar to the second case, except for the fact the h^2-θ plot shows an increase in h^2 with respect to θ while still maintaining local convexity at the transition.

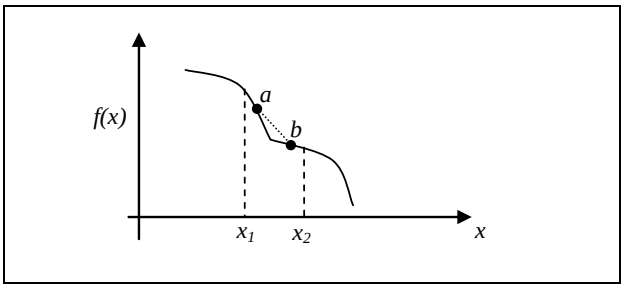


Figure 13. Illustration of local convexity of a function within an interval (x_1, x_2). The line-segment joining *a* and *b* (shown dotted) lies entirely above *f(x)*.

The behavior of the transition in the neighborhoods indicated within dotted circles in Fig. 12 can be verified by simple trigonometric calculations based on the figures shown in the top row of Fig. 12. It is worth noting that as θ increases in each of these transitions, there is a jump from a lower value for the left derivative of the function to a higher value for the right derivative of the function. Thus situations such as the ones depicted in Fig. 10(c) are avoided, and the following fact is established.

Lemma 2a: The objective function is convex at all transitions of the elliptic arcs in the composite ellipse over the domain of interest for the constrained least-squares fitting of lines in a plane.

The arguments for the convexity of the objective function across the elliptic arcs can be extended to the convexity of the objective function across the ellipsoidal patches in the 3D cases. Following the arguments presented in [14, 15] it can be shown that the equivalent of Eq. (11) in 3D cases is the following 'parallel plane theorem'

$$v^{\mathrm{T}} M_p^{\mathrm{T}} M_p v = v^{\mathrm{T}} M_g^{\mathrm{T}} M_g v + A(|v^{\mathrm{T}}(p-g)|)^2, \qquad (12)$$

where *A* is the total area of the outer convex hull. To establish local convexity of the objective function in Eq. (12) at a transition from one elliptic patch to an adjoining elliptic patch, consider any two supporting planes P_1 and P_2 each through one of any two adjacent vertices of the outer convex hull. These two planes will intersect at a line *l*. Now, transform the coordinate system so that the *z*-axis is aligned with *l*. Then the projections onto the new *xy*-plane of outer convex hull, the centroid, and the supporting planes P_1 and P_2 will yield case analyses that are identical to those shown in Fig. 12 thus establishing the local convexity of the objective function in that projected view. Since this true for *any* two supporting planes each through one of *any*

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

two adjacent vertices of the outer convex hull, the following fact is established.

Lemma 2b: The objective function is convex at all transitions of the elliptic patches in the composite ellipsoid over the domain of interest for the constrained least-squares fitting of planes in space.

The following theorems for convexity are then obtained by combining Lemmas 1a, 1b, 1c, 2a, and 2b.

Theorem 6a: The objective function is convex in the domain of interest for the constrained least-squares fitting of lines in a plane.

Theorem 6b: The objective function is convex in the domain of interest for the constrained least-squares fitting of planes in space.

7.4 Convexity for Pairs of Lines and Pairs of Planes

Consider the cases of wedges and angular slots shown in Fig. 8. Since the each of the intersecting lines or intersecting planes is fitted independently, the convexity proofs for objective functions provided thus far apply directly to these intersecting cases.

The cases of slabs and slots need some additional argument. Fitting parallel planes and parallel lines requires the consideration of the optimization problems posed in Eqs. (3) and (4). Note that the objective function in each of these two equations is the sum of two objective functions, each of which has been shown to be convex by Theorems 6a and 6b. Also, note that the two parallel lines or parallel planes will have equal and opposing normals due to parallelism. So, with simple linear transformation, the sum of the two functions in Eq. (3) or Eq. (4) can be posed as the sum of two convex functions over the same domain (say, involving longitude θ and latitude φ in the 3D case, and just one angle θ in the 2D case). Figure 14 illustrates this using a 2D example previously seen in Fig. 9. Since the sum of convex functions that share the same domain is convex, the following facts are established.

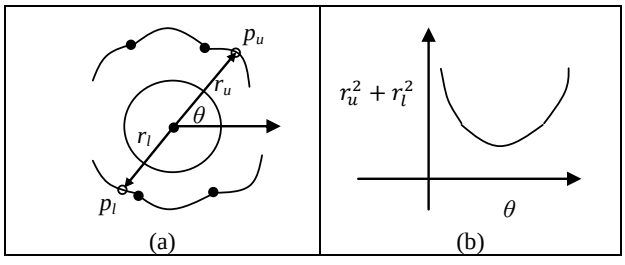


Figure 14. (a) An example of a composite ellipse in 2D, and (b) plot of the objective function $r_u^2 + r_l^2$ as a function of just one angle θ.

Theorem 7a: The objective function is convex in the domain of interest for the constrained least-squares fitting of two parallel lines in a plane.

Theorem 7b: The objective function is convex in the domain of interest for the constrained least-squares fitting of two parallel planes in space.

8. Summary and Concluding Remarks

This paper addressed the problem of establishing planar datums and systems of planar datums using constrained least-squares fitting to input points sampled on single planar features, wedges, angular slots, slabs and parallel slots. Combinatorial characterizations of the optimality conditions for these constrained optimization problems were provided with rigorous proofs. In addition, convexity of the objective functions over the domain of interest was proved.

These optimality conditions, in the form of the minimum number of points of contact, enable one to verify if a given solution is a feasible solution. In addition, the theoretical arguments that proved these optimality conditions and the proofs of convexity of the objective functions will inspire the design of new and efficient algorithms to find global optimum solutions, as demonstrated earlier using research software [14, 15]. Algorithms that exploit the results of this paper may also employ GPUs (Graphics Processing Units) to speed up the search. Further research is necessary to investigate such hardware-assisted algorithmic issues.

The constrained least-squares fitting has gained considerable support in the standards bodies owing to its theoretical advantage (as a combination of mechanical stability and numerical stability) and practical effectiveness (in software implementation). Robust implementations of constrained least-squares fitting algorithms in commercial software will be important for their acceptance by industry. Preliminary response from leading CMS vendors indicates that they are indeed beginning to implement and test their software using the results of this paper.

Acknowledgment and Disclaimer

The authors thank ISO and ASME standards experts whose advice and suggestions were invaluable in initiating and sustaining this research investigation. Any mention of commercial products or systems in this article is for information only; it does not imply recommendation or endorsement by NIST.

References

[1] ASME Y14.5-2009, Dimensioning and Tolerancing, The American Society of Mechanical Engineers, New York, 2009.

[2] ASME Y14.5.1-1994, Mathematical Definition of Dimensioning and Tolerancing Principles, The American Society of Mechanical Engineers, New York, 1994.

[3] ISO 1101:2017, Geometrical product specifications (GPS) – Geometrical tolerancing – Tolerances of form,

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

orientation, location and run-out, International Organization for Standardization, Geneva, 2017.
[4] ISO 5459:2011, Geometrical product specifications (GPS) – Geometrical tolerancing – Datums and datum systems, International Organization for Standardization, Geneva, 2011.
[5] Nielsen, H.S., The ISO Geometrical Product Specifications Handbook – Find your way in GPS, International Organization for Standardization, Geneva, 2012.
[6] Krulikowski, A., Fundamentals of Geometric Dimensioning and Tolerancing, 3rd Edition, Effective Training Inc., Westland, MI, 2012.
[7] Srinivasan, V., Geometrical product specification, in CIRP Encyclopedia of Production Engineering, Laperrière, L. and Reinhart, G. (Eds), Springer, 2014.
[8] Wick, C. and Veilleux, R.F., Tool and Manufacturing Engineers Handbook, Vol. IV Quality control and assembly, 4th Edition, Society of Manufacturing Engineers, Dearborn, MI, 1987.
[9] Golub, G. H. and Van Loan, C. F., An analysis of the total least squares problem, SIAM Journal of Numerical Analysis, Vol. 17, No. 6, pp. 883-893, December 1980.
[10] Van Huffel, S. and Vandewalle, J., The Total Least Squares Problem, SIAM Frontiers in Applied Mathematics, Philadelphia, 1991.
[11] Nievergelt, Y., Total least squares: State-of-the-art regression in numerical analysis, SIAM Reivew, Vol.36, No. 2, pp. 258-264, June 1994.
[12] Srinivasan, V., Shakarji, C.M. and Morse, E.P., On the enduring appeal of least-squares fitting in computational coordinate metrology, ASME Journal of Computing and Information Science in Engineering, Vol. 12, March 2012.
[13] Shakarji, C.M. and Srinivasan, V., Theory and algorithms for weighted total least-squares fitting of lines, planes, and parallel planes to support tolerancing standards, ASME Journal of Computing and Information Science in Engineering, 13(3), 2013.
[14] Shakarji, C.M. and Srinivasan, V., A constrained L_2 based algorithm for standardized planar datum establishment, ASME IMECE2015-50654, Proceedings of the ASME 2015 International Mechanical Engineering Congress and Exposition, Houston, TX, Nov. 13-19, 2015.
[15] Shakarji, C.M. and Srinivasan, V., Theory and algorithm for planar datum establishment using constrained total least-squares, 14th CIRP Conference on Computer Aided Tolerancing, Gothenburg, Sweden, 2016.
[16] O'Rourke, J., Computational Geometry in C, 2nd Edition, Cambridge University Press, 1998.
[17] Shakarji, C.M. and Srinivasan, V., Datum planes based on a constrained L_1 norm, ASME Journal of Computing and Information Science in Engineering, Vol. 15, December 2015.
[18] Butler, B.P., Forbes, A.B., and Harris, P.M., Algorithms for geometric tolerance assessment, NPL Report DITC 228/94, National Physical Laboratory, U.K., 1994.
[19] Anthony, G.T. et al., Chebyshev Best-Fit Geometric Elements, NPL Report DITC 221/93, National Physical Laboratory, U.K., 1993.
[20] Schoenherr, S., Quadratic Programming in Geometric Optimization: Theory, Implementation, and Applications, Doctoral Thesis, ETH Zurich, 2002.
[21] Golub, G. H., and Van Loan, C. F., Matrix Computations, 3rd Edition, The Johns Hopkins University Press, Baltimore, 1996.
[22] Gill, P.E., Murray, W., and Wright, M.H., Practical Optimization, Emerald Group Publishing Ltd., 1982.
[23] Trefethen, L. N., and Bau III, D., Numerical Linear Algebra, SIAM, Philadelphia, 1997.

Shakarji, Craig; Srinivasan, Vijay.
"Convexity and Optimality Conditions for Constrained Least-Squares Fitting of Planes and Parallel Planes to Establish Datums."
Paper presented at The ASME 2017 International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9, 2017.

MECA SENS 2017 Materials Research Forum LLC
Materials Research Proceedings **4** (2018) 129-134 doi: http://dx.doi.org/10.21741/9781945291678-20

Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-Ray Diffraction and Slitting Method

V. Luzin[1,a*], K. Spencer[2,b], M.R. Hill[3,c], T. Wei[1,d], M. Law[1,e] and T. Gnäupel-Herold[4,f]

[1]Australian Nuclear Science & Technology Organisation, Lucas Heights, NSW 2234, Australia

[2]School of Mechanical and Mining Engineering, The University of Queensland, QLD 4072, Australia

[3]Department of Mechanical and Aerospace Engineering, University of California, Davis, CA 95616, USA

[4]NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

[a]vladimir.luzin@ansto.gov.au, [b]krspencer@gmail.com, [c]mrhill@ucdavis.edu, [d]tao.wei@ansto.gov.au, [e]michael.law@ ansto.gov.au, [f]thomas.gnaeupel-herold@nist.gov

Keywords: Residual Stress, Coatings, Cold Spray, Neutron Diffraction

Abstract. The residual stress profiles in Cu and Al coatings sprayed using kinetic metallization to thickness of ~2 mm have been studied. Due to specific parameters of the cold-spray process and particular combination of materials, coatings and substrates, the residual stresses are low with magnitudes of the order of a few tens of MPa. This poses challenges on accuracy and resolution when measuring through-thickness stress distributions. Three experimental techniques - neutron diffraction, X-ray diffraction and a slitting method - were used to measure through-thickness stress distributions in the substrate-coating systems. All three techniques demonstrated acceptable accuracy and resolutions suitable for analyzing stress profiles. Advantages and disadvantages of each technique are discussed.

Introduction

Coatings of many different materials and thicknesses find their use in various surface enhancement applications such as wear resistance, corrosion protection, insulation, etc. They can be deposited on surfaces of engineering components by a number of techniques including thermal and cold spray. These techniques are energetic processes that generally induce residual stresses either through thermal effects or kinetic impact. The residual stresses can be detrimental for the coating's mechanical integrity or functional performance, therefore stress control or mitigation is usually required.

For coatings with thickness of a few millimeters, several stress measurement techniques can be used. Neutron diffraction stress measurement has proven to be a useful method for thick [1] and thin coatings [2] owing to some advantages: It is non-destructive and it can provide the required high resolution (down to 0.2 mm); No special sample preparation (e.g. cutting and polishing, as for X-rays) is required: Measurements can be done in a reasonable time (minutes per datum) and with high accuracy (uncertainty can be better than 5 MPa). The slitting method, despite it being a destructive technique, has also been proven to be an efficient method for stress measurements due to the high spatial resolution and accuracy it provides [3]. The laboratory X-ray diffraction technique is not commonly used for through-thickness stress measurement in coatings, nor is it associated with high spatial resolution as required for the investigation of

Gnaupel-Herold, Thomas; Hill, Michael R.; Law, Michael; Luzin, Vladimir; Spencer, Kevin; Wei, Tao.
"Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-ray Diffraction and Slitting Method."
Paper presented at 9th INTERNATIONAL CONFERENCE ON MECHANICAL STRESS EVALUATION BY NEUTRON and SYNCHROTRON RADIATION, Kruger National Park, South Africa. September 19, 2017 - September 21, 2017.

Materials Research Proceedings **4** (2018) 129-134 doi: http://dx.doi.org/10.21741/9781945291678-20

coatings. Some advances in the technique [4] and demonstration of its ability in application to thin (~1 mm) metal sheets [5] open opportunities for a new application of this technique. Since the laboratory X-ray technique is most accessible, it gives certain advantages to this technique.

In this work we report on an experimental study of the residual stress analysis in cold-spray coatings made with the three techniques mentioned.

Sample production

Two coatings were sprayed using the kinetic metallization (KM) cold-spray technique with a simple convergent barrel nozzle at a nozzle standoff distance of 12 mm and a nozzle traverse speed of 50 mm/s. The gas temperature-pressure conditions, optimized for deposition efficiency and reported in Table 1 were also used to estimate the exit velocity of the particles.

The deposition of powder on flat copper coupons (30 × 30 mm^2, 3 mm thickness) resulted in coatings with a thickness of ≈ 2.1 mm for both materials.

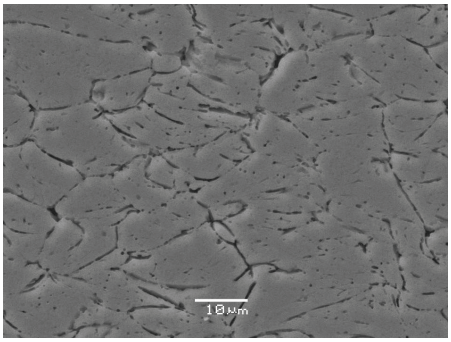


Fig. 1: Microstructure of the Al coating.

Table 1. Sample spraying conditions and parameters.

Powder Material	Average Particle Size [µm]	Driving Pressure [kPa]	Nozzle Temperature [°C]	Powder Feed Rate [g/min]	Estimated Particle Exit Velocity [m/s]
Al	15	620	140	15	585
Cu	6	620	200	15	645

Sample characterization

(i) For measurements of the Young's modulus, rectangular specimens were extracted from the bulk of the coating with approximate dimensions 30 × 5 × 2 mm^3. The Young's modulus E was determined using the impulse excitation technique (ASTM standard E1876) through acoustic measurements of the normal frequency and the sample dimensions. The accuracy of this method was better than 1 %. The same samples were used for density ρ measurements (Table 2).

(ii) In order to determine the phase composition, the full neutron diffraction patterns were measured (using the very same bar samples used for the Young's modulus measurement) in the 2θ range of 10° – 160° at a wavelength of 1.62 Å using high-resolution powder diffractometer ECHIDNA at the ANSTO OPAL research reactor [6]. The volume fractions of the phases were extracted through a refinement procedure (Table 2).

Table 2. Coating's material characterization: Young's modulus, density and phase composition.

ID	E [GPa]	% of bulk	ρ [g/mm^3]	% of bulk	Phase composition, vol %
Al	49.4 ± 0.2	69.6 ± 0.3	2.528 ± 0.016	93.6 ± 0.6	Al 100 %
Cu	104.0 ± 0.5	83.9 ± 0.6	8.670 ± 0.044	98.5 ± 0.5	Cu95 %+Cu_2O 5 %

Neutron residual stress measurements

The neutron diffraction residual stresses measurements were carried out at the NIST Center for Neutron Research using the BT8 residual stress diffractometer [7]. In correlation to the sample thicknesses a 0.5 × 0.5 × 18 mm^3 sized gauge volume was used. For the most optimal definition of the gauge volume a 90˚ ($2\theta_B$ ~ 90˚) diffraction geometry was employed by setting the take-off angle $2\theta_M$ of the Si(311) monochromator and the wavelength λ accordingly (Table 3). Through-

Gnaupel-Herold, Thomas; Hill, Michael R.; Law, Michael; Luzin, Vladimir; Spencer, Kevin; Wei, Tao.
"Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-ray Diffraction and Slitting Method."
Paper presented at 9th INTERNATIONAL CONFERENCE ON MECHANICAL STRESS EVALUATION BY NEUTRON and SYNCHROTRON RADIATION, Kruger National Park, South Africa. September 19, 2017 - September 21, 2017.

Materials Research Proceedings **4** (2018) 129-134 doi: http://dx.doi.org/10.21741/9781945291678-20

thickness measurements were done in locations within the substrate and coating with 0.3 mm spacing between points. For each measurement point d-spacings were measured in the two principal directions, normal to the surface and in-plane. The d-spacings of the (311) reflections were measured with sufficient statistics to provide at least 5×10^{-5} uncertainty in the strain. Since Cu is a stronger neutron scatterer, 10 minutes per measurement point (t) were adequate for Cu, while 90 minutes were required for Al. Assuming a balanced biaxial plane-stress state, the stress values were calculated according to the procedure [1] from the measured d-spacings and the diffraction elastic constants estimated by the self-consistent Kröner method [9] (Table 3).

Table 3. Neutron instrument setting and material constants for the measured reflections.

	d [Å]	$2\theta_M$	λ [Å]	$2\theta_B$	S_1 [TPa^{-1}]	½S_2 [TPa^{-1}]	t [min]
Cu(311)	1.10	70.0˚	1.55	89.7˚	-3.38	12.58	10
Al(311)	1.22	79.8˚	1.72	89.5˚	-5.16	19.57	90

X-ray diffraction measurements

On completion of the neutron diffraction measurements, the samples were cut in halves and the surface of the cross-sectional cut investigated with X-ray measurements (after appropriate surface preparation). The measurement technique used in the high spatial resolution measurements was described previously [4]. It utilizes a very narrow vertical beam, 0.05 mm in our case, a combination of Ω- and Ψ-angles, as well as sample rotation applied in such a way that ensures that the X-ray beam projection remains parallel to the surface/interface, therefore representing the same through-thickness depth, and does not exceed the desired spatial resolution, 0.1 mm in our case.

The Cu-tube K-α radiation was used to measure the Cu(420) and Al(511/333) reflections. Although measurement time with such a small beam is an order of magnitude longer (t = 120 seconds for each orientation), many through-thickness locations could be measured in the available beamtime. The experimental details of the X-ray measurements are summarised in Table 4.

The primary interest was on locations immediately adjacent to the interface, thus posing a challenge for the neutron technique. Thus, 6 points in each coating were measured in steps of 0.1 mm with the first point located 0.1 mm away from the interface. The average uncertainty of X-ray stress values was better than 5 MPa, a prerequisite to resolve subtle stress variations.

An attempt was also made to measure stresses in the Cu substrate, but due to the large grains, statistical variations were too great to provide a reliable result notwithstanding measurements taken for two Cu reflections, Cu(420) and Cu(331) and used in combination. Therefore, this data was omitted from the publication.

Table 4. X-ray instrument setting and material constants for the measured reflections.

	d [Å]	Tube	λ [Å]	$2\theta_B$	S_1 [TPa^{-1}]	½S_2 [TPa^{-1}]	t [sec]
Cu (420)	0.8061	Cu	1.54433	145˚	-2.87	11.01	120
Al (511/333)	0.7793	Cu	1.54433	163˚	-5.15	19.69	120

Slitting measurements

For the slitting method stress measurements [3], 5 mm thick slices were cut from each of the X-ray samples to have $5 \times 5 \times 30$ mm^3 bars available for these tests. (Through-thickness dimension size of 5.3 mm is composed of the 3.2 mm thick Cu substrate and a 2.1 mm thick coating). A strain gauge attached to the surface opposite to the coated surface was used to measure strain changes when a slit was cut through the sample thickness by wire EDM. Cut depth increments

Gnaupel-Herold, Thomas; Hill, Michael R.; Law, Michael; Luzin, Vladimir; Spencer, Kevin; Wei, Tao.
"Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-ray Diffraction and Slitting Method."
Paper presented at 9th INTERNATIONAL CONFERENCE ON MECHANICAL STRESS EVALUATION BY NEUTRON and SYNCHROTRON RADIATION, Kruger National Park, South Africa. September 19, 2017 - September 21, 2017.

were 0.1 - 0.2 mm, with a final cut depth of about 4.5 mm from the top surface. This allowed for reconstruction of the stresses in the coatings and most of the substrate thicknesses.

The reconstruction of the stress profiles from the strain gauge data was done in two different ways: The unit-pulse analysis approach using a Tikhonov regularisation to reconstruct stress profiles of different degrees of smoothness and statistical robustness; Eigenstrain analysis approach where the stress profile solution is sought within *a priori* determined class of functions (e.g. polynomial of n^{th} order) as a best fit to the experimental data. A comparison of the two approaches is given if Fig. 2 for the Cu/Cu system. The second approach (eigenstrain, order = 1) in the case of sprayed coating has certain advantages since it is known that: (i) the stress profile in the substrate is a linear function (elastic bending response to bring about force and bending moment balances); (ii) the stress profile in the coating typically is a very smooth function, very close to a linear function (as demonstrated below by modelling results). In fact, this type of solution is the only one that can provide a reasonable result in the most difficult case, i.e. the Al/Cu system.

FEM modelling

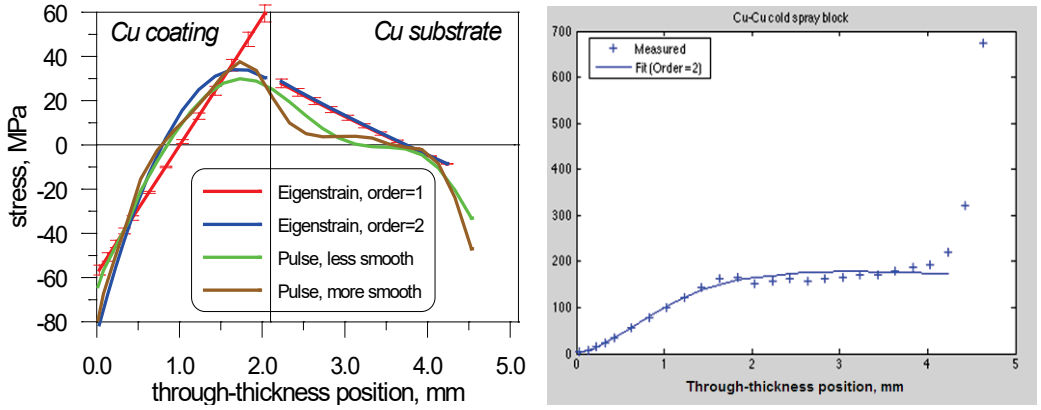


Fig. 2: Slitting method analysis of the Cu/Cu system: (left) several solutions of the reconstructed stress and (right) a corresponding strain function (µstrain vertical axis) fitting the strain gauge experimental data for one of the solutions. The results are provided in the coordinate system of the EDM cut: zero point corresponds to the top coating surface.

The most critical part of the stress analysis in coatings is the geometrical conditions of the investigated samples and the relationship between measured stresses. In reality, the measured strains/stresses are not exactly the same (Fig. 3): (i) The neutron measurements were done in the central part of the sample, away from the edges, ensuring that the original equi-biaxial (typical for most coatings) stress state is measured; (ii) For the X-ray technique measurements are made on a cut, so the stress state is not equi-biaxial anymore and the normal stress component is eliminated by the presence of a free surface. One in-plane component remains, but may be altered by the cut; (iii) For convenience of the slitting method, the original sample was altered even further. The second cut to produce a 5 mm bar further reduces constraints from the other parts of the half-sample, thus changing the stress state again.

To address the issue of the different sample geometries and relationships between stress states the Chill modelling approach [9] was applied on a generalized bi-metal two-layer plate sample to emulate the coating/substrate system. The system was 5 mm thick Al (E = 70 GPa) and 3 mm thick steel (E = 210 GPa) with plate dimensions 100 × 100 mm^2. Instead of a stress profile

Gnaupel-Herold, Thomas; Hill, Michael R.; Law, Michael; Luzin, Vladimir; Spencer, Kevin; Wei, Tao.
"Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-ray Diffraction and Slitting Method."
Paper presented at 9th INTERNATIONAL CONFERENCE ON MECHANICAL STRESS EVALUATION BY NEUTRON and SYNCHROTRON RADIATION, Kruger National Park, South Africa. September 19, 2017 - September 21, 2017.

generated by a deposition process, a stress profile was induced here thermally through a combination of different coefficients of thermal expansion and temperature increments.

Using ABAQUS, the two consecutive cuts were simulated by tracing the stress relaxations and stress profiles at each step. The results of the simulation are given in Fig. 4. Although some geometry effects are visible in the simulation as distortions of the stress profile from the ideal (theoretically linear), the general trend of the stress distribution demonstrates that a stress relaxation effect is observable, but not at a significant level. Some 80 – 90 % of the original stress magnitude is preserved in the bar or edge sample. Although, some corrections (a factor) can be applied to account for the partial stress relaxation of the remaining in-plane component, but in our case this effect was considered negligible.

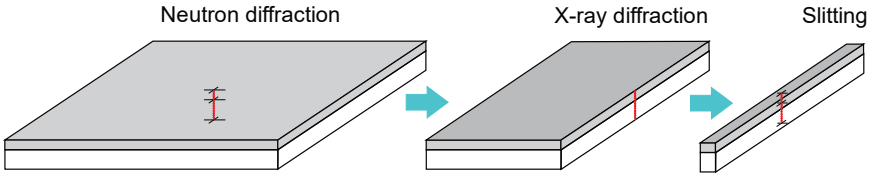


Fig. 3: Three different sample geometries used for the investigation with the three techniques.

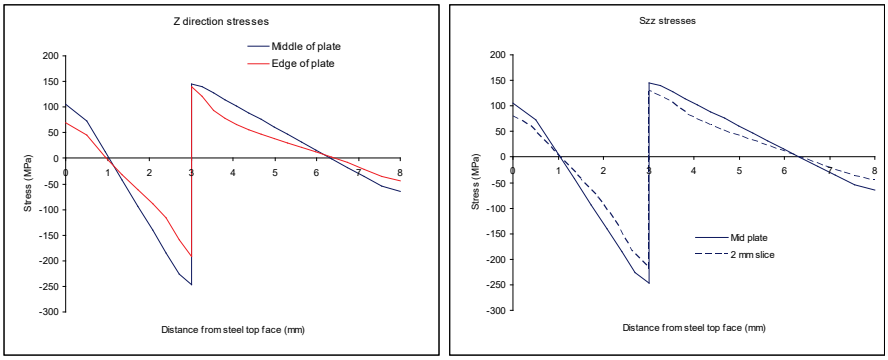


Fig. 4: Relaxation of the stress on the sample edge (left) and in the thin bar (right) in comparison with the stress in the middle of the uncut sample.

Results and discussion

The results of the three experimental methods are combined in Fig. 5. Overall, all three methods demonstrated an ability to resolve the stresses with good accuracy, < 5 – 10 MPa, even in the case of very subtle stress distribution in the Al/Cu system. It is therefore confirmed that any of the techniques is suitable for stress measurements in coatings.

While neutrons and X-rays are in very good agreement, the slitting methods results have some visible deviation from the diffraction results, as is seen most vividly in the Cu/Cu system. Upon reviewing, there are at least two objective reasons for results of the slitting method to differ. Both of them are linked to the fact that the top surface is not smooth, but exhibits a significant variation of ±0.4 mm from the average, as measured from minimum and maximum values (Fig. 5). Due to this, depending on the exact location of the EDM cut, the effective coating thickness can vary within a range, and strain relaxation readings will be sensitive to this effect. Secondly, for the stress reconstruction procedure, the coating and substrate thicknesses are parameters of the elastic model and therefore, if average (vs local) thickness is used, this will also impact the final stress values.

Gnaupel-Herold, Thomas; Hill, Michael R.; Law, Michael; Luzin, Vladimir; Spencer, Kevin; Wei, Tao.
"Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-ray Diffraction and Slitting Method."
Paper presented at 9th INTERNATIONAL CONFERENCE ON MECHANICAL STRESS EVALUATION BY NEUTRON and SYNCHROTRON RADIATION, Kruger National Park, South Africa. September 19, 2017 - September 21, 2017.

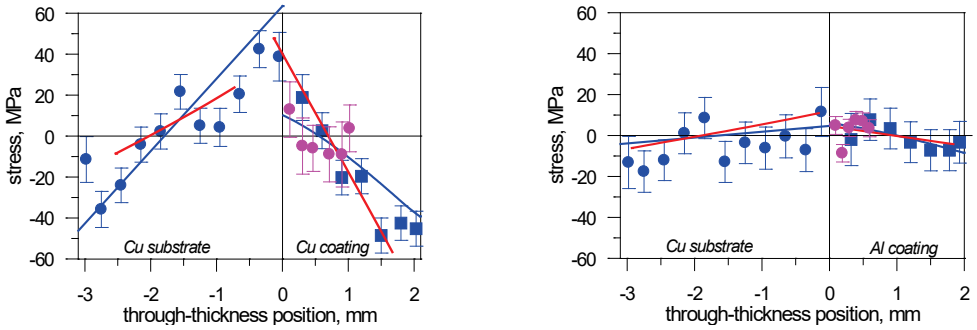


Fig. 5: The results of stress determination by the three methods for the two coating systems, Cu/Cu (left) and Al/Cu (right). Neutron data are shown in blue symbols with Tsui & Clyne model [10] (blue line) fitting to the datasets. X-ray data are shown as magenta symbols. The red line is for the slitting method results.

In the cases of neutrons and X-rays, the measurement and stress calculation procedures rely on average parameters leading to more robust results with better agreement.

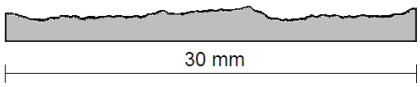


Fig. 6: Variation of the Al coating thickness profile: t_{min} = 1.6 mm, t_{max} = 2.8 mm, t_{ave} = 2.0 mm.

References

[1] V. Luzin, A. Valarezo and S. Sampath, Through-thickness Residual Stress Measurement in Metal and Ceramic Spray Coatings by Neutron Diffraction, Mater. Sci. Forum 571-572 (2008) 315-320. https://doi.org/10.4028/www.scientific.net/MSF.571-572.315

[2] V. Luzin, A. Vackel, A. Valarezo and S. Sampath, Neutron Through-Thickness Stress Measurements in Coatings with High Spatial Resolution, Mater. Sci. Forum 905 (2017) 165-173. https://doi.org/10.4028/www.scientific.net/MSF.905.165

[3] M.R. Hill, The Slitting Method, in: Practical Residual Stress Measurement Methods, John Wiley & Sons, Ltd, 2013, pp. 89-108. https://doi.org/10.1002/9781118402832.ch4

[4] T. Gnaupel-Herold, Formalism for the determination of intermediate stress gradients using X-ray diffraction, J. Appl. Crystallogr. 42 (2009) 192-197. https://doi.org/10.1107/S0021889809004300

[5] T. Gnaeupel-Herold, T. Foecke, M. Iadicola and S. Banovic, in: SAE International, 2005.

[6] K.-D. Liss, B. Hunter, M. Hagen, T. Noakes and S. Kennedy, Echidna – the new high-resolution powder diffractometer being built at OPAL, Physica B: Condensed Matter 385–386 Part 2 (2006) 1010-1012. https://doi.org/10.1016/j.physb.2006.05.322

[7] http://www.ncnr.nist.gov/instruments/darts/

[8] T. Gnaupel-Herold, P.C. Brand and H.J. Prask, Calculation of Single-Crystal Elastic Constants for Cubic Crystal Symmetry from Powder Diffraction Data, J. Appl. Crystallogr. 31 (1998) 929-935. https://doi.org/10.1107/S002188989800898X

[9] M. Law, O. Kirstein and V. Luzin, An assessment of the effect of cutting welded samples on residual stress measurements by chill modelling, J. Strain Anal. Eng. Des. 45 (2010) 567-573. https://doi.org/10.1177/030932471004500807

[10] Y.C. Tsui and T.W. Clyne, An analytical model for predicting residual stresses in progressively deposited coatings Part 1: Planar geometry, Thin Solid Films 306 (1997) 23-33. https://doi.org/10.1016/S0040-6090(97)00199-5

Gnaupel-Herold, Thomas; Hill, Michael R.; Law, Michael; Luzin, Vladimir; Spencer, Kevin; Wei, Tao.
"Stress Profiling in Cold-Spray Coatings by Different Experimental Techniques: Neutron Diffraction, X-ray Diffraction and Slitting Method."
Paper presented at 9th INTERNATIONAL CONFERENCE ON MECHANICAL STRESS EVALUATION BY NEUTRON and SYNCHROTRON RADIATION, Kruger National Park, South Africa. September 19, 2017 - September 21, 2017.

Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy

Martin Y. Sohn*[a], Dong Ryoung Lee[a], Bryan M. Barnes[a], Ronald Dixson[a], Richard M. Silver[a], Sang-Soo Choi[b]

[a]Engineering Physics Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD, USA 20899-8212; [b]PKL-Photronics, 493-3 Sungsung, Cheonan, Choongnam, South Korea 330-300

ABSTRACT

A critical challenge in optical critical dimension metrology, that requires high measurement sensitivity as well as high throughput, is the dimensional measurements of features sized below the optical resolution limit. This paper investigates the relationships among dimensional sensitivity and key illumination beam conditions (e.g., angular illumination, partial coherence) for photomask feature characterization. Scatterfield images at the edge areas of multiple line structures on a Molybdenum Silicide (MoSi) photomask are analyzed to establish sensitivity to dimensional changes. Actinic scatterfield imaging experiments for these features are performed using the NIST 193 nm Scatterfield Microscope, designed to enable engineered illumination beams at the target. Illumination configurations that improve sensitivity are identified from imaging edges of multiple line targets having linewidths and spaces of about 1/3 wavelength.

Keywords: Optical critical dimension metrology, scatterfield imaging microscopy, measurement sensitivity, partial coherence factor, illumination engineering.

1. INTRODUCTION

Achieving higher densities and smaller dimensions in semiconductor manufacturing technology requires continuous advances in measurement technology. Compared to other metrology techniques, optical metrology techniques have been important to the industry because of their nondestructive character, higher speed, and relatively lower cost [1-3]. A key challenge for optical methods is accurate dimensional measurement of ever-decreasing deep subwavelength feature sizes. Characterizing such features requires high measurement sensitivity with low uncertainties as well as high throughput [4].

Optical scatterfield imaging microscopy techniques have enabled the characterization of nanoscale features sized well below the resolution limit with high sensitivity by analyzing far field images formed by light scattered from the features [5-7]. Engineering of the illumination beam is essential for yielding high sensitivity in scatterfield imaging as the characteristics of the scattered light depend not only on the parameters of the features but also strongly on the incident light conditions such as incident angle, angular intensity, polarization and partial coherence.

The National Institute of Standards and Technology (NIST) has developed reflection scatterfield imaging microscopes operating at visible and deep ultra-violet (DUV) wavelengths for the characterization of nanoscale features on both wafers and photomasks. Model-based CD measurements for finite grating structures of sub-20 nm features have been reported using a visible light scatterfield imaging microscope by Fourier domain normalization method for parametric simulations for comparison with experimental results, yielding sub-nanometer uncertainties [8]. The measured intensity profiles showed high ringing effects at the edges of 100-line gratings and interference effects between both edges of 30-line gratings as a small numerical aperture illumination was imposed on the targets. Notable, a low partial coherence factor may have triggered these edge effects and might be related to the sensitivity of the parametric measurements using the scatterfield imaging microscopy technique. From these results, we noticed that both partial coherence, expressed as the illumination numerical aperture (NA) relative to the collection numerical aperture [9], and image quality have been important for high resolution imaging microscopy and for lithography, thus the relationship between the partial coherence and measurement sensitivity for image-based CD metrology warranted further investigation.

* martin.sohn@nist.gov

Updated 3/20/14

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

In this paper, we present a dimensional measurement sensitivity analysis with respect to various illumination conditions as a function of the partial coherence factor using the NIST 193 nm scatterfield imaging microscope for multiple line gratings on a MoSi photomask that have linewidths of nominally 66 nm to 72 nm with a 2 nm increment. Methodology, details for experiments, and sensitivity measurement results are presented with optimal illumination configurations for improving sensitivity.

2. METHODOLOGY FOR DIMENSIONAL METROLOGY

Conventional imaging microscopy has a relatively wide field of view for imaging the targets, though the imaging resolution is limited due to the diffraction limit. Scatterometry is widely used to measure the sub-wavelength dimensions of nanoscale features using scattered intensity distributions containing dimensional information. Using a variety of incidence angles and wavelengths, measured intensities are fit to models generated based on electromagnetic simulation to determine geometrical parameters [10]. Scatterfield imaging microscopy can combine the angular illumination capability of the scatterometry technique and the imaging capability of the conventional imaging microscopy into one platform as shown in Fig. 1. This combination allows manipulation of the illumination angles and shapes that can help more information from nanoscale features through capturing images and potentially varying the focus. By using this method, the three-dimensional scattered fields at the sample are tailored to yield scattered far field images with high sensitivity for improved measurements.

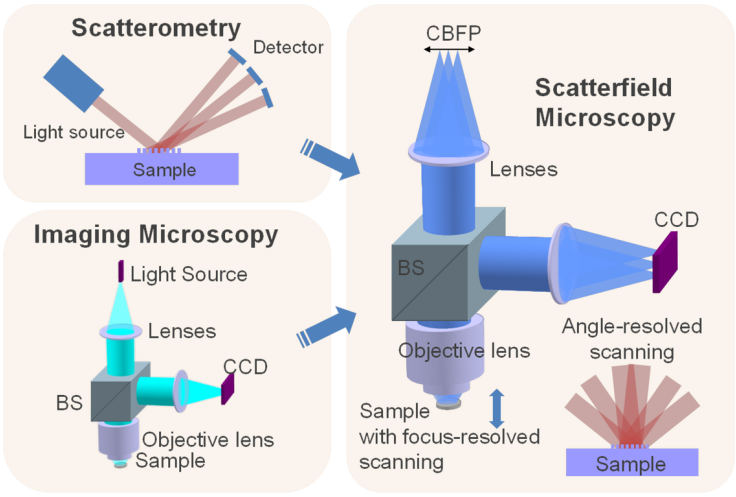


Fig. 1. Schematic of angle-resolved scatterfield imaging microscopy with focus variation.

Tailoring the fields scattered from the sample in scatterfield imaging microscopy is performed by controlling the illumination shape at a conjugate plane of the back focal plane (CBFP) of the objective lens according to the Köhler illumination principle. Fig. 2 shows a schematic of illumination engineering for scatterfield imaging that can be performed at the CBFP. Assuming that the optical system has a Köhler configuration for illumination, a diverging beam from a position at the CBFP illuminates the sample plane as a collimated beam at a corresponding angle. Based on this relationship between the lateral position and the illumination angle, the illumination beam shape can be manipulated by placing apertures designed for specific angles of incidence or by scanning a finite aperture.

This capability has been applied recently to the characterization of tool functions for quantitative nanoscale feature measurement based on parametric modelling analysis. Tool functions for characterizing the illumination and collection path deviations and errors are obtained as intensity functions of illumination angle by scanning a finite aperture at the CBFP [11]. These tool functions are ultimately used to adjust simulated scattered fields for improved fitting to obtain nanoscale feature parameters such as linewidth, height, or sidewall angle based on the captured far field scattered images at the image plane. Using a visible light scatterfield imaging microscope (λ= 450 nm), NIST researchers performed model-based CD measurements for finite grating structures of sub-20 nm silicon-on-silicon lines utilizing a Fourier normalization method which utilizes these tool functions [8]. Scattered field images captured at the image plane were compared against simulation data that were normalized by the instrument characterization and parametric fitting with

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

sub-nanometer uncertainties. Observing the profiles of these nanoscale features, a small illumination numerical aperture leads to a high ringing at the edges of a relatively large target (100-line grating) and an interference between fields scattered by both edges of a small targets (30-line finite grating). These edge effects may make the scatterfield imaging measurements sensitive to linewidth variations. A low partial coherence factor (0.13/0.95) imposed on the illumination of the optical system triggered these edge effects and could be related to the sensitivity of the parametric measurements using the scatterfield imaging microscopy technique, thus this work investigates the interplay among the edge effects, dimensional sensitivity, and partial coherence in a scatterfield imaging system.

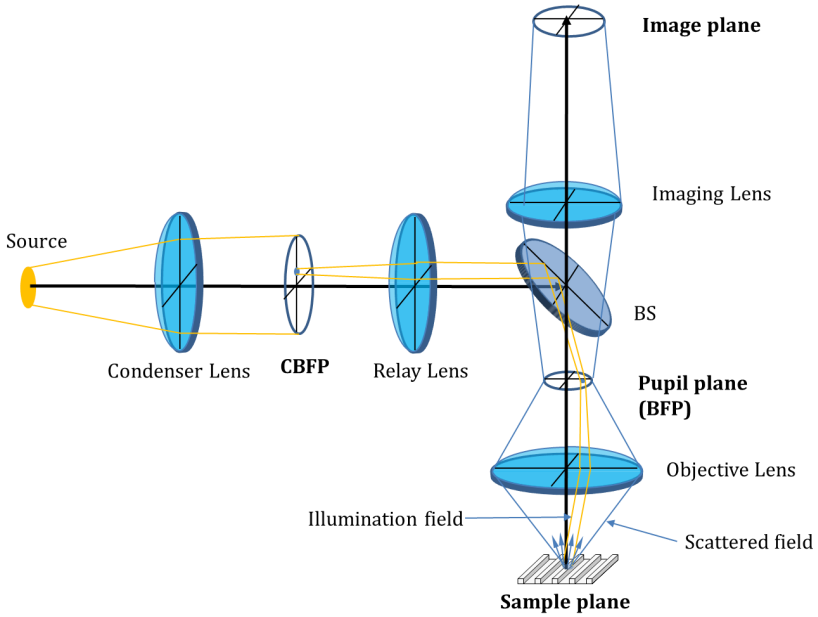
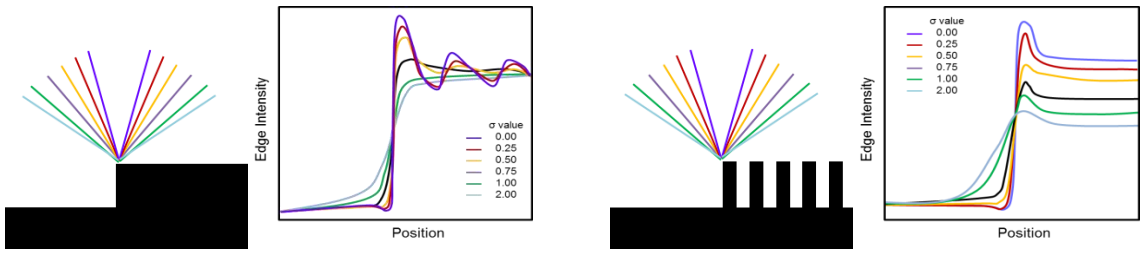


Fig. 2. Illumination engineering schematic for scatterfield imaging using CBFP.

The partial coherence factor $\sigma = NA_{ill} / NA_{col}$, as a metric signifying illumination partial coherence degree in the optical imaging system, is long-established to be important for improving image quality in imaging microscopy system and in lithography exposure process [12-16]. Fig. 3 (a) shows variations in the intensity profiles at a step edge as a function of partial coherence factor [17]. A lower partial coherence factor signifies higher partial coherence generating more ringing effects at the edge. If the edge of a nanoscale grating is considered, the intensity profiles as a function of the partial coherence factor will consist of overshoot at the edge and flat intensity variations at the grating area, as shown in Fig. 3(b). This phenomenon may allow the estimation of measurement sensitivity for the gratings with various linewidths and partial coherence of the imaging system.



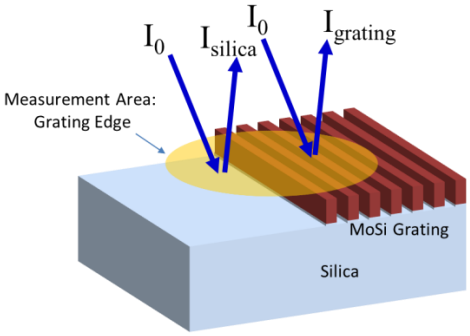
(a) At a step edge

(b) At the edge of multiple lines

Fig. 3. Schematic of hypothetical scattered intensity profiles at the edge for two types of steps.

In this work, measurement of the line and space structure is improved by imaging of the edge area, including both the grating and the substrate. Scatterfield imaging microscopy has an advantage that the imaging area has a relatively large

field of view, which can cover the substrate and grating areas simultaneously as shown in Fig. 4. These capabilities for positioning and spatial selectivity within an image both are useful for measuring the relative signal intensity variations from different linewidth gratings. The normalized signal intensity of the grating area is obtained as the ratio of intensities scattered from the grating and reflected from the substrate. By comparing these relative intensities for various grating linewidths using the substrate as a reference, measurement sensitivity analysis can be performed without modeling. The scattered intensity from the grating varies with the grating's geometric parameters such as linewidth and pitch, with dominant 0-th order scattering away from the edge and a continuum of scattered orders from the grating area. These variations can be used to estimate the sensitivity as a function of partial coherence.



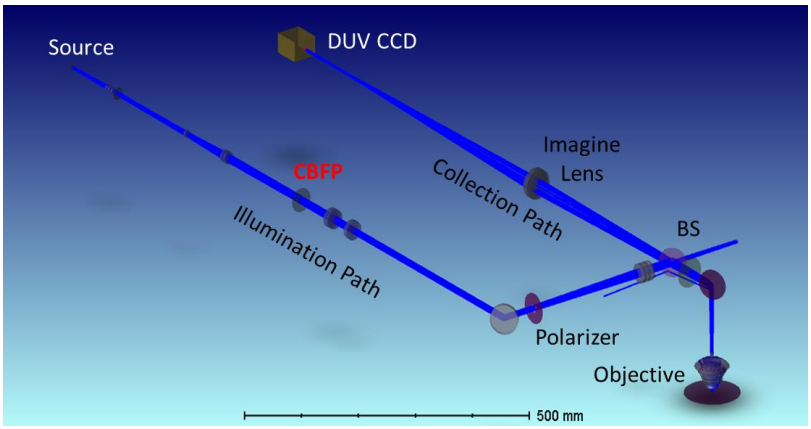
$$I_{norm} = \frac{I_{grating}}{I_{silica}}$$

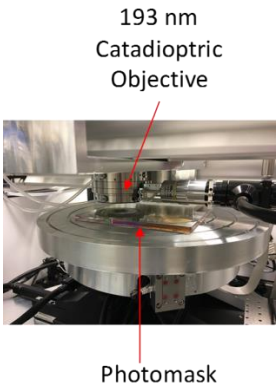
$$I_{silica} = r_{silica} I_0$$

Fig. 4. Edge imaging method for differentiating scattered intensities from grating area.

3. EXPERIMENTATION

The NIST 193 nm scatterfield imaging microscope system uses an ArF excimer laser as a source, which is the actinic source used for the deep ultraviolet (DUV) immersion lithography exposure process. The microscope adopts a catadioptric objective lens that has an NA of 0.13-0.74, which has a central reflection mirror that blocks light within a numerical aperture of 0.13. The instrument is designed to have a relatively large telecentric CBFP of a diameter of 11.6 mm for enhancing the controlling of the illumination shape as shown in Fig. 5 (a). Various apertures for illumination engineering are placed at the CBFP.





(a) Optics designed (b) Photomask on the stage

Fig. 5. The NIST 193 nm scatterfield imaging microscope platform.

To engineer the illumination in terms of partial coherence using this DUV scatterfield imaging microscope, three kinds of illuminations are created by three basic aperture shapes: horizontally rectangular, vertically rectangular, and circular apertures, as shown in Fig. 6. Partial coherence can be altered by varying the size of these apertures. For horizontally rectangular apertures, the y-direction illumination angle is both narrow and constant, while the x-direction angle is

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

varied by changing the length D_x. For vertically rectangular apertures, the y-direction angle covers the full NA, but the x-direction angle is varied by changing the length D_x. For circular apertures, all radial angles are varied with aperture diameter D_x. These aperture size variations are designed to effectively test the interactions of the shaped illumination beams with the grating line for both directions.

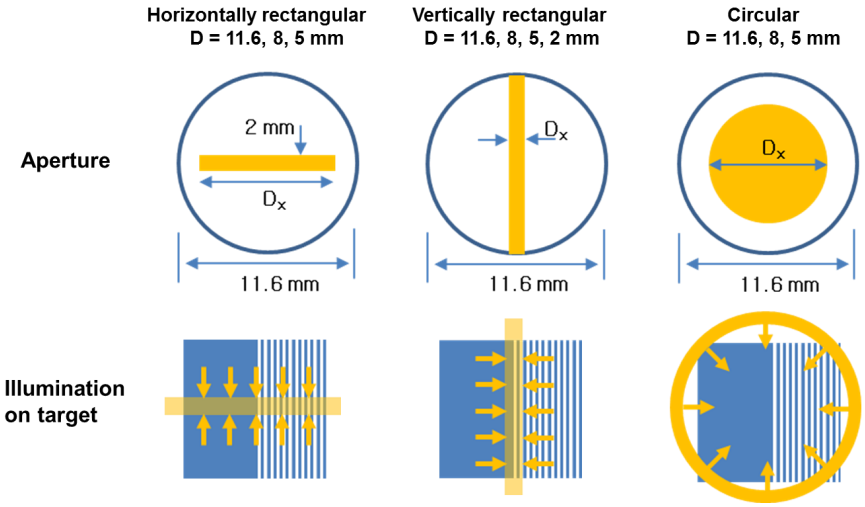


Fig. 6. Apertures for engineering the illumination by varying the partial coherence.

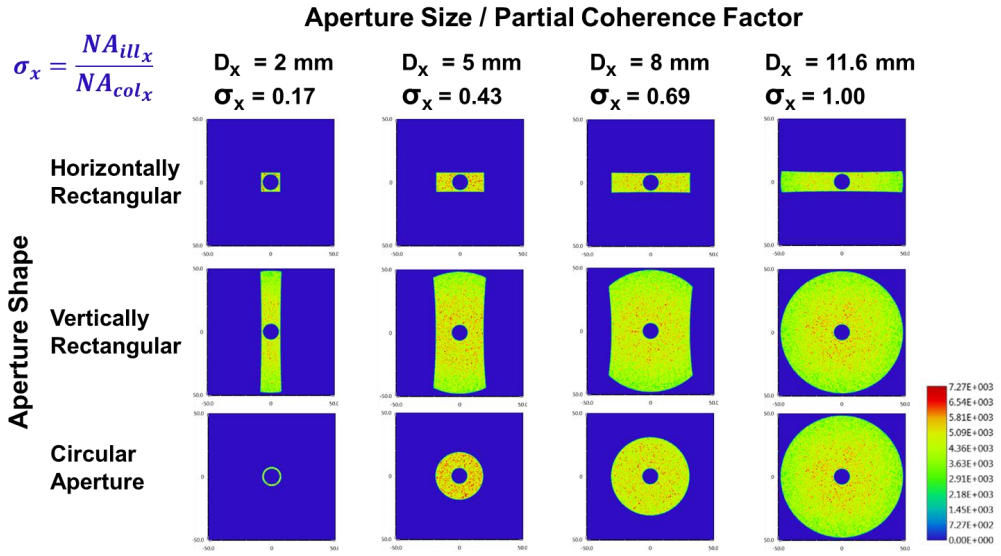
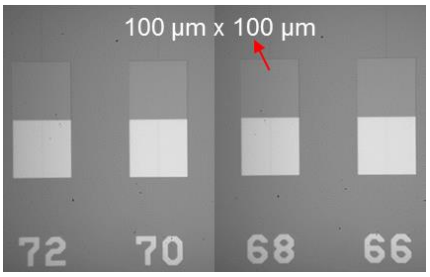


Fig. 7. Simulated angular illumination with respect to the partial coherence factors σ_x.

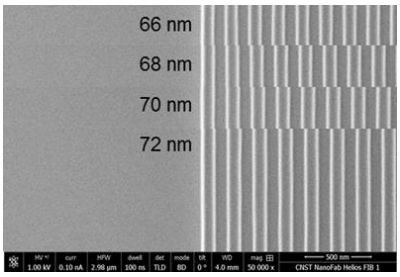
Using these apertures, angular illumination conditions have been simulated with the optics designed as shown in Fig. 7. The picture array shows the angular illumination shapes at the sample plane in the angle domain as a function of the partial coherence factor for values between 0.17 and 1.0. The partial coherence factor σ_x is varied from 0.17 to 1. The central obscurations are observed due to the central mirror reflection inside of the catadioptric objective lens. The illuminations at a partial coherence factor of 0.17 for the horizontally rectangular and circular apertures yield very small

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

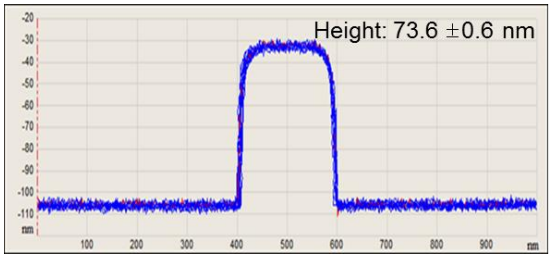
amounts of energy for illumination, which leads to low signal to noise ratios on the scattered intensity profiles, and therefore, are excluded from the experiments.



(a) Wide-view optical microscope images



(b) Focused Ion Beam images

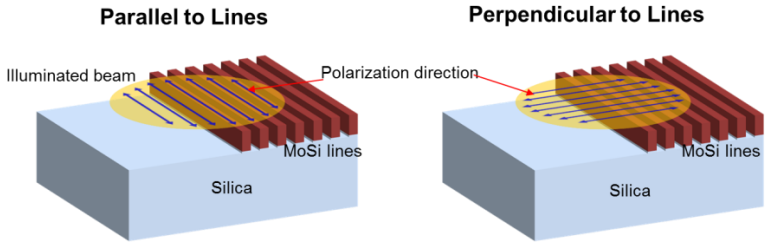


(c) AFM profile of a line extension

Nominal linewidth (nm)	Measured linewidth (nm)	Standard deviation (nm)
66	63.49	1.17
68	65.59	1.17
70	67.96	1.17
72	69.89	1.17

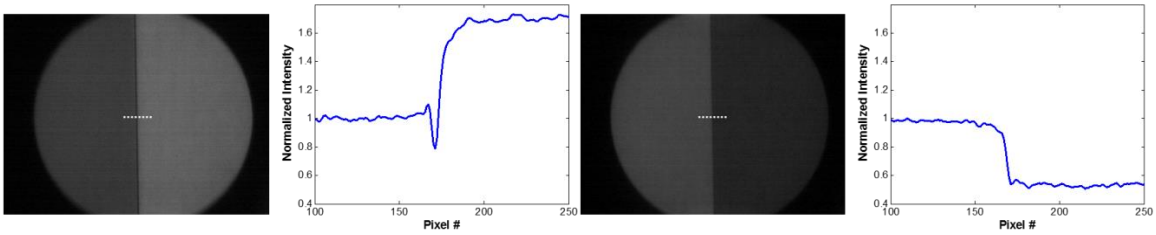
(d) Linewidth values measured with AFM

Fig. 8. Target images captured by various metrology tools.



(a) Polarization parallel to lines

(b) Polarization perpendicular to lines



(c) Reflected intensity for parallel polarization

(d) Reflected intensity for perpendicular polarization

Fig. 9. Dependence of edge intensity profiles on illumination polarization.

Sensitivity measurements have been performed for MoSi line gratings on a silica substrate with designed linewidths of 66 nm to 72 nm with 2 nm variation and a 1:1 line/space ratio. Such photomask targets are widely used for lithographic imaging with the phase shift mask technique [18,19]. Fig. 8 shows images taken with a visible-light microscope with a

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

wide field of view, a focused ion beam (FIB) microscope, and an atomic force microscope (AFM) as reference measurements. The line grating area is 100 μm x 100 μm and each grating has line extension at the center to allow linewidth measurement using AFM as shown in Fig. 8 (a). Gradual variations for each linewidth are shown as an overlapped figure in Fig. 8 (b). The line heights are measured at 73.6 ± 0.6 nm with varied actual linewidths between 63.5 and 69.9 nm as shown in Fig. 8 (c) and (d).

Because the scattered light from gratings is sensitive to the polarization of the incident beam, edge profiles have been measured using two illumination polarizations. Fig. 9 shows the scattered intensity profiles at the edge area for the two orthogonally polarized illumination beams. The reflected intensities between the substrate (left) and the grating (right) change with respect to polarization direction: the intensity at the substrate is lower than at the grating for parallel polarization, while the intensity at the substrate is higher than at the grating for perpendicular polarization, as shown in Fig. 9 (c) and (d). The intensity changes over the silica area are due to the different reflectivity distribution for incident beam angle with different polarizations, which can be calculated from the Fresnel equation. But the grating area intensity changes are due to the scattering interactions among the lines and with the silica substrate.

4. RESULTS AND DISCUSSION

To better measure intensity differentiations among sets of line gratings, intensity profiles were collected for both focused and defocused positions by moving the sample along the z axis as shown in Fig. 10. Each column shows a set of different focus intensity profiles of the 4 different gratings shown in the FIB images inset, varying the target z-position over a range of 2 μm with 1 μm steps and varying the partial coherence factor from 0.17 to 1. A set of 4 intensity distributions for 4 gratings are used to calculate the sensitivity at a specific focus and partial coherence. Edge signal shapes follow trends that are typical of partial coherence variation at a step edge, following the trends shown in Fig. 3 (b). It is observed that the stronger signal intensity variations among the gratings occur at the edges with increased overshoot effects as the partial coherence factor decreases, which signifies increasing partial coherence. Also, the normalized intensity difference between the substrate and the grating areas become larger as the partial coherence increases.

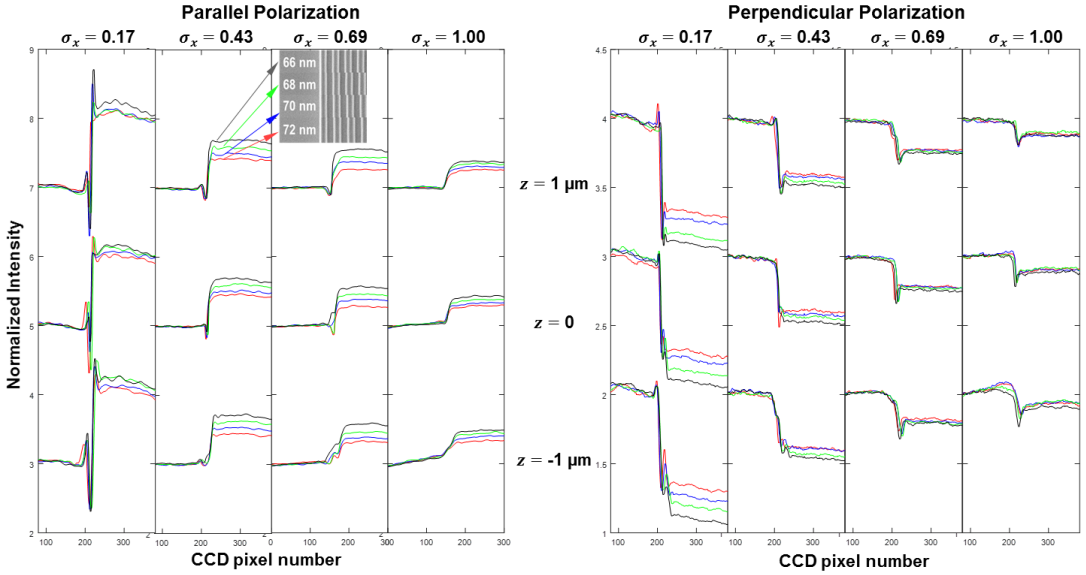


Fig. 10. Edge intensity signals along the grating linewidth and pitch variations in z position variation.

From these intensity distribution variations, empirical sensitivities are calculated for the different illumination instances as shown in Fig. 7 apart from model-based sensitivity analysis [20]. The sensitivity, μ is defined as ratio between the measurand difference Δl and the signal to noise ratio δs for two targets to be measured [21]. The measurands in this analysis are the linewidth values measured by AFM and the signal to noise ratio values are calculated using the mean

intensity difference $\overline{I_2} - \overline{I_1}$ and the standard deviation averages which are obtained from repeated measurements. Note, the 1:1 correspondence of the line and spaces ties linewidth to pitch.

$$\mu = \frac{\Delta l}{\delta s}, \qquad \text{where } \delta s = \frac{abs(\bar{I}_2 - \bar{I}_1)}{\sqrt{\frac{1}{2}\left(std_{I_2}^2 + std_{I_1}^2\right)}}$$

The grating structures for sensitivity measurements either have varied linewidths with a fixed pitch or varied linewidths with varied pitches. For the former, the intensity signal varies only due to scattered field distributions from line geometries as the angles of diffraction orders are fixed. For the latter, the intensity signal varies as field distributions scattered by both line geometries and varied angle of high order diffraction beam due to pitch. In this sensitivity analysis, we used various grating structures, each with its own linewidth and pitch. While linewidth and line geometry may influence the optical response, the intensity variations in this paper are presumed to be due to the pitch variations. And thus the sensitivity analysis is based on the latter case.

Using this empirical sensitivity concept, the sensitivities as a function of the partial coherence factor are obtained from the variations in the normalized intensity distributions as shown in Fig. 11. The graphs summarize the mean intensity results for all investigated cases of apertures, vertically rectangular, horizontally rectangular, and circular. Each point is mean value for repeated measurements and the error bar for each point is the standard deviation. Separations and standard deviations on each graph lead to the sensitivity variations. The vertical aperture shows higher signal variations that are acceptable. These scattered intensity data from the three aperture shape designs, two polarizations, various partial coherence factors, have been processed to yield sensitivity data for three focus positions: at the substrate and one micrometer above and below the substrate.

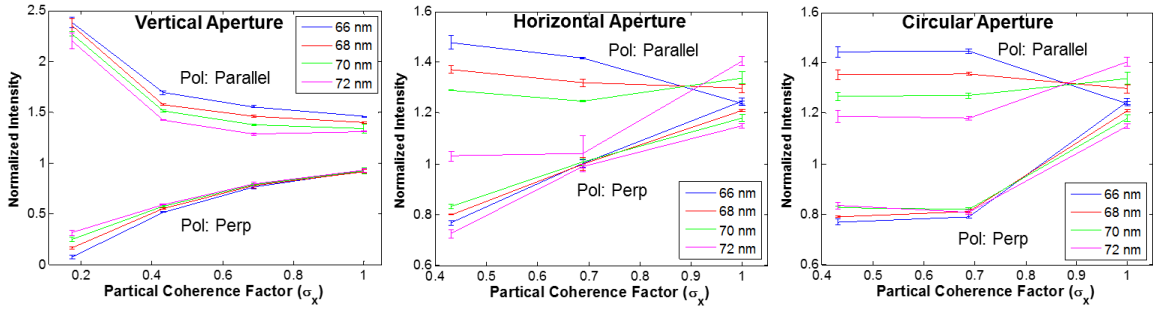


Fig. 11. Normalized scattered intensity signals along the partial coherence factor: each point is mean value for repeated measurements and the error bar for each point is the standard deviation.

Good (< 0.3 nm) Sufficient (0.3 - 0.5 nm)

Sample Position in Z			-1 µm				0 µm				+1 µm			
Partial Coherence Factor (σ_x)			0.17	0.43	0.69	1	0.17	0.43	0.69	1	0.17	0.43	0.69	1
Vertical Aperture	Parallel Pol	Mean	3.055	0.343	0.369	1.222	2.724	0.272	0.249	1.332	2.96	0.441	0.234	1.252
		Std	1.627	0.17	0.189	0.898	2.142	0.087	0.028	1.069	1.662	0.096	0.086	1.254
	Perpendicular Pol	Mean	0.493	0.704	3.935	13.07	0.52	0.494	4.862	13.69	0.372	0.205	3.044	7.514
		Std	0.226	0.386	0.994	10.34	0.197	0.243	4.027	6.056	0.215	0.049	1.137	3.295
Horizontal Aperture	Parallel Pol	Mean		0.543	0.459	1.097		0.283	0.351	0.721		0.391	0.498	0.694
		Std		0.303	0.141	0.626		0.164	0.121	0.353		0.217	0.113	0.404
	Perpendicular Pol	Mean		0.617	16.71	0.77		0.34	9.133	0.669		0.641	6.579	1.17
		Std		0.383	23.63	0.275		0.087	9.108	0.17		0.429	1.671	0.411
Annular Aperture	Parallel Pol	Mean		0.501	0.369	1.097		0.48	0.232	0.721		0.372	0.229	0.694
		Std		0.242	0.192	0.626		0.021	0.036	0.353		0.127	0.092	0.404
	Perpendicular Pol	Mean		1.192	1.459	0.77		1.228	1.23	0.669		1.121	5.593	1.17
		Std		0.618	1.025	0.275		1.138	0.626	0.17		0.221	6.505	0.411

Fig. 12. Sensitivity estimation for partial coherence factor

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

Fig. 12 shows a metric for determining empirical sensitivity that is applied for each aperture case, polarization, and linewidth. Some of the entries are marked as yellow and green, indicating a sensitivity between 0.3 nm and 0.5 nm and sensitivities less than 0.3 nm, respectively. Agreeing with the graphs in Fig. 11, the illuminations using vertical aperture yield better sensitivities more frequently than the other cases. From these data, in general the optimal illumination conditions for scatterfield imaging measurements consists of a vertical aperture, parallel polarization, and medium partial coherence factor range of about $\sigma = 0.4 \sim 0.7$. Note, the best observed sensitivity, however, is found at 1 micrometer above the substrate using perpendicular polarization.

5. CONCLUSIONS

We investigated empirical sensitivity for a set of MoSi photomask targets with respect to partial coherence using DUV scatterfield imaging microscopy at a wavelength of 193 nm, concluding that partial coherence impacts dimensional measurement sensitivity for scatterfield metrology. The higher partial coherence results in stronger signal intensities at the grating area and improved sensitivities for these targets could be found by tailoring the partial coherence. In this experiment, the optimal illumination condition consists of a vertical rectangle with polarization parallel to the grating lines and a partial coherence factor range between 0.4 ~ 0.7. Optimized measurement sensitivity for scatterfield imaging microscopy is less than 0.3 nm with uncertainties of 0.1 nm or better. Additional target designs should be investigated, analyzing linewidth sensitivity with respect to partial coherence using targets with fixed pitch and varying linewidths. Based on this empirical sensitivity study, it is projected that the DUV scatterfield imaging microscopy technique is feasible to measure nanoscale critical dimensions with sub-nanometer sensitivity, required for advanced CD measurements based on electromagnetic parametric model.

REFERENCES

[1] J. M. Shalf and R. Leland, "Computing beyond Moore's Law," Computer **48**(12), 14-23 (2015).

[2] M. Asano, R. Yoshikawa, T. Hirano, H. Abe, K. Matsuki, H. Tsuda, M. Komori, T. Ojima, H. Yonemitsu, A. Kawamoto, "Metrology and inspection required for next generation lithography," Jpn. J. Appl. Phys. **56**, 06GA01 (2017).

[3] L. Sun, T. Kohyama, K. Takeda, H. Nozawa, Y. Asakawa, T. Kagalwala, G. Lobb, F. Mont, X. Dai, S. Pal, W. Wang, J. Kye, F. Goodwin, "High throughput and dense sampling metrology for process control," Proc. SPIE **10145**, 101452D (2017).

[4] R. M. Silver, T. Germer, R. Attota, B. M. Barnes, B. Bunday, J. Allgair, E. Marx, J. Jun, "Fundamental limits of optical critical dimension metrology: a simulation study," Proc. SPIE **6518**, 65180U (2007).

[5] R. M. Silver, B. M. Barnes, R. Attota, J. S. Jun, M. T. Stocker, E. Mark, H. J. Patrick, "Extending the limits of image-based optical metrology," Appl. Opt. **46**, 4248-4257 (2007).

[6] B. M. Barnes, R. Attota, R. Quintanilha, M. Y. Sohn, R. M. Silver, "Characterizing a scatterfield optical platform for semiconductor metrology," Meas. Sci. Technol. **22**, 024003 (2010).

[7] B. M. Barnes, R. Quintanilha, M. Y. Sohn, H. Zhou, R. M. Silver, "Optical illumination optimization for patterned defect inspection," Proc. SPIE **7971**, 79710D-1 (2011).

[8] J. Qin, R. M. Silver, B. M. Barnes, H. Zhou, R. Dixson, M-A. Henn, "Deep subwavelength nanometric image reconstruction using Fourier domain optical normalization," Light-Sci. Appl. **5**, e16038 (2016).

[9] H. H. Hopkins, "On the diffraction theory of optical images," Proc. Royal Soc. London Series A, **217**(1130) 408-432 (1953).

[10] C. J. Raymond, *et. al.*, "Multiparameter grating metrology using optical scatterometry," J. Vac. Sci. Technol. B **15**, 361 (1997).

[11] J. Qin, R. M. Silver, B. M. Barnes, H. Zhou, F. Goasmat, "Fourier domain optical tool normalization for quantitative parametric image reconstruction," Appl. Opt. **52**(26) 6512-6522 (2013).

[12] E. C. Kintner and R. M. Sillitto, "Edge-ringing in partially coherent imaging," Opt. Acta **24**(5), 591-605 (1977).

[13] M. Singh, H. Lajunen, J. Tervo, J. Turunen, "Imaging with partially coherent light: elementary-field approach," Opt. Express **23**(22), 244084 (2015).

[14] B. Smith, "The saga of sigma: Influences of illumination throughout optical generations," Proc. SPIE **9052**, 905204 (2008).

Barnes, Bryan; Choi, Sang-Soo; Dixson, Ronald; Lee, Dong; Silver, Richard; Sohn, Martin.
"Dimensional measurement sensitivity analysis for a MoSi photomask using DUV reflection scatterfield imaging microscopy."
Paper presented at SPIE Photomask Technology and EUV Lithography, Monterey, CA, United States. September 11, 2017 - September 14, 2017.

[15] X. Ma and G. Arce, “Binary mask optimization for inverse lithography with partially coherent illumination,” J. Opt. Soc. Am. A **25**, 2960 – 2970 (2008).
[16] P. Naulleau, *et. al.*, “Sub-70 nm extreme ultraviolet lithography at the advanced light source static microfield exposure station using the engineering test stand set-2 optic,” J. Vac. Sci. Technol. B **20**(6), 2829 – 2833 (2002).
[17] S. M. Shakeri, L. J. Vliet, S. Stallinga, “Impact of partial coherence on the apparent optical transfer function derived from the response to amplitude edges,” Appl. Opt. **56**(12), 3518 (2017).
[18] W. Ahn, et. al., “Development of high-transmittance phase-shifting mask for ArF immersion lithography,” Proc. SPIE **9658**, 965808 (2015).
[19] S. Zhang, et. al., “Performance comparison between attenuated PSM and Opaque MoSi on Glass (OMOG) mask in sub-32nm Litho Process,” ECS Trans. **44**(1), 249-256 (2012).
[20] P. Vagos, J. Hu, Z. Liu, S. Rabello, “Uncertainty and sensitivity analysis and its applications in OCD measurements,” Proc. SPIE **7272**, 72721N (2009).
[21] M. Stocker, B. M. Barnes, M. Sohn, E. Stanfield, R. M. Silver, "Development of large aperture projection scatterometry for catalyst loading evaluation in proton exchange membrane fuel cells," J. Power Sources **364**, 130-137 (2017).

Influence of Lucky Defect Distributions on Early TDDB Failures in SiC Power MOSFETs

J. Chbili[1,2*], Z. Chbili[3], A. Matsuda[4], K. P. Cheung[1], J. T. Ryan[1], J. P. Campbell[1], M. Lahbabi[2]
[1]Engineering Physics Division, NIST, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[2]Laboratoire SSC, Faculté des Sciences et Techniques, USMBA B.P. 2202 Fez, Morocco
[3]GLOBALFOUNDRIES Inc., 400 Stonebreak Road Extension, Malta, NY, 12020, USA
[4]National Institute for Material Science, Ibaraki 305-0047, Japan.
*email : jaafar.chbili@nist.gov, phone : 301-975-0462

***Abstract*— In this work, we explore the effect of different defect profiles on the occurrence of early time-dependent-dielectric-breakdown (TDDB) to forecast the defect profile present in commercial grade SiC/SiO_2 DMOSFETs. Early failure simulations are performed using the recently developed "lucky defect" model. The model shows that the bulk defects in the gate oxide are the likely culprit for early TDDB failures through an increase in tunneling current via trap-assisted-tunneling (TAT). We show that an exponential distribution of "lucky defects" in the oxide bulk affects the failure distribution in a similar fashion to what we observe experimentally. We also identify the implications of under-sampling the population in these extrinsically dominated failure distributions. Armed with these tools, we show that, the speculated carbon rich transition layer at or near the interface is not likely present in the measured DMOSFETs.**

Keywords—Reliability Testing; SiC; TDDB; Extrinsic Failures; Lucky Defect Model;

I. INTRODUCTION

Silicon Carbide (SiC) DMOSFET production is reaching a maturity level that allows it to replace conventional silicon devices in the power market [1, 2]. Recent studies have shown that the bias temperature instability (BTI) risk for such devices is diminishing due to recent advances in annealing and interface passivation [2, 3]. The intrinsic TDDB performance was also shown to be comparable, if not better than, that of similar SiO_2/Si devices [4]. However, the transition to large volume production requires better control of extrinsic defects and a low level of early failures which is what dictates the overall product reliability. Unfortunately, extrinsic reliability in SiO_2/SiC has received little to no attention. This study greatly extends earlier efforts [5] to identify the origin of these early failures by establishing a correlation between the extrinsic tails of measured failure distributions and user defined lucky defect profiles for arbitrarily chosen device populations.

II. EARLY FAILURES IN SILICON CARBIDE DEVICES

Figure 1 shows the collective TDDB failure distribution collected from over 430 SiC DMOSFETs with 50 nm thick thermal oxide SiO_2 at different fields and different temperatures. Typically, SiC DMOSFET TDDB distributions suffer from a tail similar to what is shown in fig. 1. This tail is frequently *ignored* to focus on the intrinsic reliability. Since the oxide and area are identical for all tested devices, and assuming the breakdown mechanism is the same for all stress fields and temperatures, all groups have the same intrinsic β but a different $T_{63\%}$. Thus, we can normalize the different failure distributions to increase the sample size. The main observation from fig.1 is a distribution tail of around 7.5%. In this paper, we will examine the possible source of such a distribution tail and its connection to the presence of traps in the oxide.

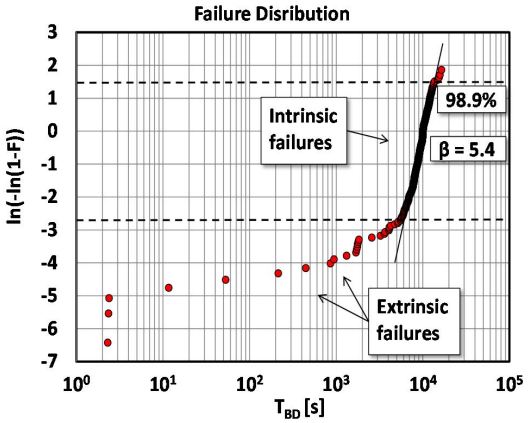


Fig. 1. TDDB results on a total of 432 SiC DMOSFET (200 µm x 200 µm) with a 50 nm thick oxide SiO_2/SiC. The collective data from several TDDB test fields (6.8 MV cm^{-1} to 10 MV cm^{-1}) and temperatures (25 °C to 275 °C) were normalized to a breakdown time of 10^4 s.

III. LUCKY DEFECT MODEL

A newly developed model has attributed the occurrence of early defects in SiO_2/SiC to the presence of "lucky defects" in the oxide bulk [5]. In this model, a "lucky" defect with the appropriate energy level and spatial location (fig. 2) will locally enhance the gate leakage current through trap-assisted-tunneling (TAT) during high-field TDDB stress (fig. 3) where the tunneling current is normally in the Fowler-Nordheim (FN) regime.

Campbell, Jason; Chbili, Jaafar; Chbili, Zakariae; Cheung, Kin; Lahbabi, M; Matsuda, Asahiko; Ryan, Jason.
"Influence of Lucky Defect Distributions on Early TDDB Failures in SiC Power MOSFETs."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

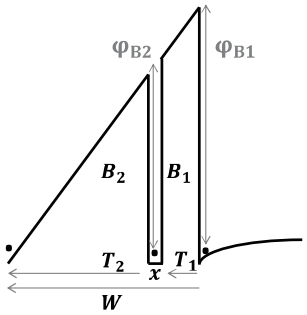


Fig. 2. TAT probability is controlled by the probability to tunnel into the defect (T1), and the probability to tunnel out to the oxide conduction band (T2). The total tunneling current resulting from TAT is at its maximum when T1 = T2.

The increased leakage current leads to an increase in oxide wear-out rate and therefore a shorter lifetime. This effect reaches its maximum when the defect is located at a "sweet spot", $x(E_{OX})$, defined as:

$$x(E_{ox}) = W(E_{ox}) * [1 - \sqrt{2}/2] \quad (1)$$

where W is the total width of the FN barrier at the stress field E_{OX}. Thus, the lifetime reduction depends on the physical location of the "lucky defect" in the dielectric and the corresponding amount of gate leakage enhancement due to TAT. Hence the reduced lifetime in the presence of a "lucky defect":

$$T_{BD}^{*} = T_{BD} \bullet \left(\frac{J_{FN}}{J_{FN}+J_{TAT}}\right)\left(\frac{A}{A_{DF}}\right)^{\frac{1}{\beta}} \quad (2)$$

where J_{FN} is the FN component of the leakage, J_{TAT} the TAT component of the leakage, A the area of the device and A_{DF} the area of the "lucky defect" assumed to be ~1 nm^2 [5].

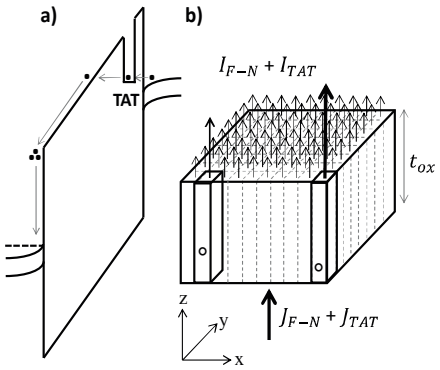


Fig. 3. Tunneling current in the presence of traps: a) traps at the "lucky energy" resulting in b) local increase in current.

Using this model, we examine the effect of different "lucky defects" profiles on the TDDB performance.

IV. MODEL IMPLICATIONS ON FAILURE DISTRIBUTIONS

The defect density of the oxide (10^8 cm^{-3}) and the different distributions considered are related to the oxide process and can be considered as a by-product of the SiO_2 oxide growth on SiC. To simplify the simulation, we assume all defects have the appropriate energy to enhance tunneling. Thus, we consider a wafer area of (2 cm x 2 cm) containing ten thousand DMOSFETs, each with the same area as the one used experimentally (200 µm x 200 µm) and the same 50 nm thick SiO_2 dielectric, as shown in fig. 4.

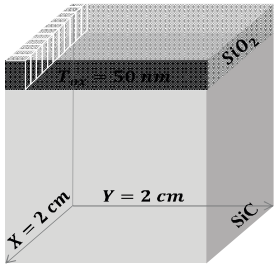


Fig. 4. Dimensions of DMOSFETS considered.

Fig. 5 shows a uniform distribution of "lucky defects" which would be similar to a uniform contamination during oxide growth.

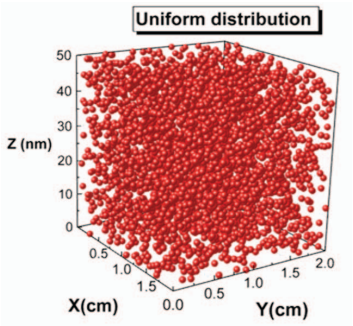


Fig. 5. Uniform lucky defect distribution with a density of defect of 10^8 cm^{-3}

As mentioned earlier, the most lifetime reduction is caused by defects that are at or close to a sweet-spot (Eq. 1), but the current enhancement decreases exponentially as the defect is placed further away from the sweet-spot (Eq. 2). Thus, for the uniform distribution of defect shown in fig. 5, only a small fraction of defects located closer to the SiO_2/SiC interface result in early fails. This is evident in fig. 6a where a tail of 1.18 % "extrinsic" fails is obtained. However, if we consider 50 randomly selected devices out of the distribution, which is a similar statistical sample to a usual wafer level TDDB test, we would not observe any early fails (fig. 6c). This early TDDB assessment would lead to conclusions that no further screening is needed. However, if we consider a sample size of 400 (fig. 6b) which is similar to high temperature operating life (HTOL) qualification tests at the package level, we would observe a few

Campbell, Jason; Chbili, Jaafar; Chbili, Zakariae; Cheung, Kin; Lahbabi, M; Matsuda, Asahiko; Ryan, Jason.
"Influence of Lucky Defect Distributions on Early TDDB Failures in SiC Power MOSFETs."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

outliers and an apparent inconsistency with the wafer-level data.

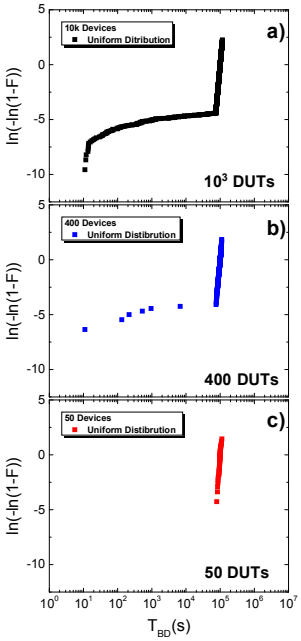


Fig. 6. Uniform distribution of lucky defects. a) shows the TDDB failure distribution of 10000 devices showing tails of early failures, b) shows the resulting TDDB distribution for 400 devices sample size which is similar to an HTOL stress, and c) shows the TDDB distribution with 50 devices sample size which is similar to a routine TDDB test at wafer level.

Fig. 7 shows a normal distribution of defects mostly contained within 10 nm of the interface. This distribution is often cited in the SiC literature as a few nanometers thick carbon rich transition layer has been observed at the SiO_2/SiC interface using electron microscopy [6,7].

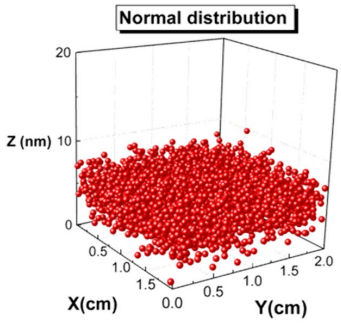


Fig. 7. Normal lucky defect distribution mostly contained within 10 nm of the interface with a density of defect of 10^8 cm^{-3}.

Similar to a uniform distribution, this defect profile only results in a 0.83 % tail which might not be captured during a wafer level test (fig. 8c).

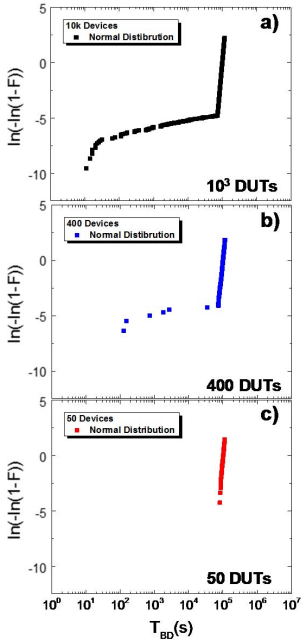


Fig. 8. Normal distribution of lucky defects mostly contained within 10 nm of the interface. a) shows the TDDB failure distribution of 10000 devices showing tails of early failures, b) shows the resulting TDDB distribution for 400 devices sample size which is similar to an HTOL stress, and c) shows the TDDB distribution with 50 devices sample size which is similar to a routine TDDB test at wafer level.

However, if we consider a narrower transition layer, only a nanometer from the interface (fig. 9), we see a large increase in early failures and the tail reaches 31.95 % of the population (fig. 10c). This tail is much higher than the 7.5 % tail that was observed experimentally (fig. 1).

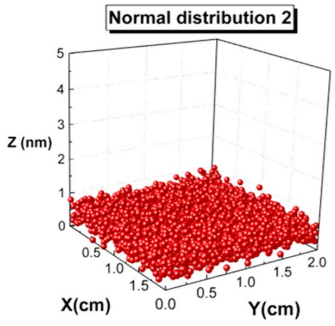


Fig. 9. Normal lucky defect distribution mostly contained within 1 nm with a density of defect of 10^8 cm^{-3}.

Campbell, Jason; Chbili, Jaafar; Chbili, Zakariae; Cheung, Kin; Lahbabi, M; Matsuda, Asahiko; Ryan, Jason.
"Influence of Lucky Defect Distributions on Early TDDB Failures in SiC Power MOSFETs."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

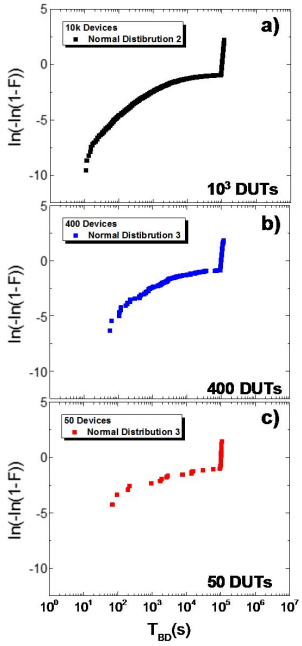


Fig. 10. Normal distribution of lucky defects mostly contained within 1 nm of the interface. a) shows the TDDB failure distribution of 10000 devices showing tails of early failures, b) shows the resulting TDDB distribution for 400 devices sample size which is similar to an HTOL stress, and c) shows the TDDB distribution with 50 devices sample size which is similar to a routine TDDB test at wafer level.

Finally, we consider an exponential distribution with the highest concentration closest to the interface (fig. 11). The resulting tail observed in (fig. 12a) is of 10.95 % which is similar to what was measured experimentally. This distribution is the most physically sound, since carbon related species are created as a byproduct of the thermal growth. At the initial phase of growth, the SiO_2 is thin enough that the carbon related species escape easily. However, as the oxide grows thicker, the carbon species cannot escape and lead to a larger concentration of defects closer to the interface.

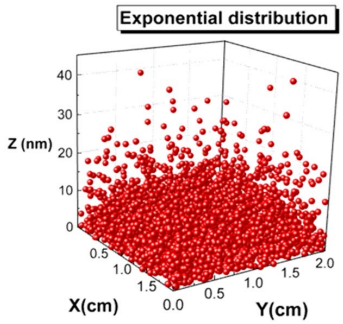


Fig. 11. Exponential lucky defect distribution with a density of defect of 10^8 cm^{-3}.

Fig. 12c shows that a 50 samples TDDB early assessment would lead to the conclusion that a moderate screening is a viable option since only 2-4 fails were observed. However, a screening criteria based on that test would not solve the early failure tail for the total population and fails would still be observed, even at HTOL.

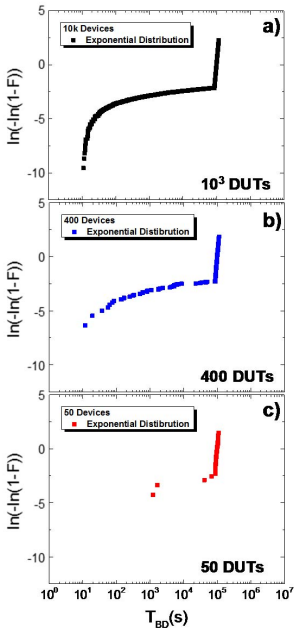


Fig. 12. Exponential distribution of lucky defects. a) shows the TDDB failure distribution of 10000 devices showing tails of early failures, b) shows the resulting TDDB distribution for 400 devices sample size which is similar to an HTOL stress, and c) shows the TDDB distribution with 50 devices sample size which is similar to a routine TDDB test at wafer level.

V. Conclusion

Despite the SiC community's efforts to improve contamination and processing, the products are still plagued with extrinsic-like fails. The observations reported here recommend a shift in focus from contamination control to oxide growth process improvements. Moreover, the simulations show that the sample size of the TDDB failure distributions is crucial to the TDDB assessment and the screening process can be destructive rendering the data meaningless.

References

[1] R. Eden, HIS Technology 2014; J. W. Palmour, IEDM 2014;
[2] D. R. Hughartet al, ECS Trans. 58, 2013;
[3] Z. Chbili et al, Mater. Sci. Forum 858, 2016;
[4] Z. Chbili et al, IEEE Trans. Electron Devices 63, 2016;
[5] K. C. Chang et al, Appl. Phys. Lett 77, 2000;
[6] T. Zheleva et al, Appl. Phys. Lett 93, 2008

Campbell, Jason; Chbili, Jaafar; Chbili, Zakariae; Cheung, Kin; Lahbabi, M; Matsuda, Asahiko; Ryan, Jason.
"Influence of Lucky Defect Distributions on Early TDDB Failures in SiC Power MOSFETs."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

Wafer Level EDMR: Magnetic Resonance in a Probing Station

Duane J. McCrory,
Engineering Science
Penn State University
University Park, USA
djm5501@psu.edu

Mark A. Anders
Materials Science
Penn State University
University Park, USA
andersmark@gmail.com

Jason T Ryan
Engineering Physics
NIST
Gaithersburg, USA
jason.ryan@nist.gov

Pragya R. Shrestha
Theiss Research, La Jolla, CA;
Engineering Physics, NIST
Gaithersburg, USA
pragya.shrestha@nist.gov

Kin P. Cheung
Engineering Physics
NIST
Gaithersburg, USA
kin.cheung@nist.gov

Patrick M. Lenahan
Engineering Science
Penn State University
University Park, USA
pmlesm@engr.psu.edu

Jason P. Campbell
Engineering Physics
NIST
Gaithersburg, USA
jason.campbell@nist.gov

***Abstract*— We report on a novel semiconductor reliability technique that merges electrically detected magnetic resonance (EDMR) with a conventional semiconductor wafer probing station. This union with a semiconductor probing station allows EDMR measurements to be performed at the wafer level. Our measurements forgo a microwave cavity or resonator for a very small non-resonant near field microwave probe [1]. Bipolar amplification effect (BAE) [4] and spin dependent charge pumping (SDCP) [5] were demonstrated on various SiC MOSFET structures. These measurements were made via frequency-swept EDMR. The elimination of the resonance cavity, and incorporation with a wafer probing station, greatly simplifies the EDMR detection scheme and offers promise for widespread EDMR adoption in semiconductor reliability laboratories.**

Keywords*— *EDMR; wafer level; defects; magnetic resonance; reliability

I. INTRODUCTION

As semiconductor technology continues to scale and newer material systems are introduced, it is imperative to have reliability measurements that are capable of identifying performance limiting defects at the atomic scale. Currently, most reliability measurements are performed using wafer probing stations. The electrical measurements made at these stations have proven widespread and high-volume applicability for counting charges in device measurements. However, these measurements yield limited information about the atomic-scale nature of the defects that capture and emit charges and ultimately govern reliability. The most sensitive technique for identifying atomic-scale defects in semiconductor devices is electrically detected magnetic resonance (EDMR). This technique, first proposed as spin dependent recombination (SDR) by Lepine in 1972 [6], is a derivative of electron paramagnetic resonance (EPR) in which the measurement is performed on fully processed device structures. EDMR is at least ten million times more sensitive than conventional EPR [7], and thus adequately addresses the sensitivity needs associated with advanced device structures. EDMR is a powerful tool for investigating performance limiting defects in many different material systems [7]-[11]. Unfortunately, EDMR measurements require significant sample preparation and bulky spectrometers. Despite the rich information provided by EDMR, the experimental barriers relegate it as a niche technique only suitable for individual device measurements. Clearly there is a need for an EDMR spectrometer that eliminates sample preparation and can be incorporated into a conventional wafer probing station. In this paper, we demonstrate the viability of a wafer level EDMR spectrometer utilizing a non-resonant near field microwave probe (Fig. 1). This approach is demonstrated via

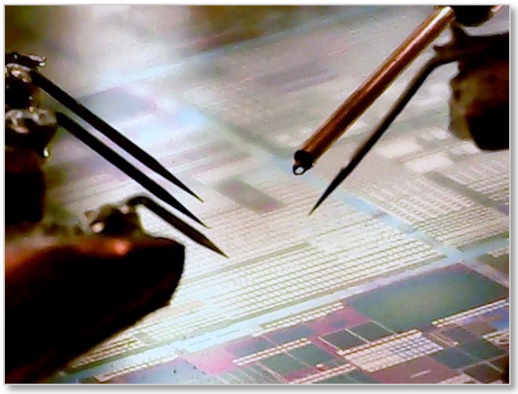


Fig. 1. Side view of the non-resonant near field microwave probe (top right) utilized in all EDMR measurements made in this paper. The four other probes are home-built high-frequency current probes used to measure device characteristics.

Anders, Mark; Campbell, Jason; Cheung, Kin; Lenahan, Patrick; McCrory, Duane; Ryan, Jason; Shrestha, Pragya.
"Wafer Level EDMR: Magnetic Resonance in a Probing Station."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

observation of Si-vacancy defects [7]-[8] which dominate the performance and reliability of SiC MOSFETs.

II. Experimental Details

All MOSFETs analyzed with the wafer-level spectrometer were 4H-SiC n-MOSFETs with varying gate sizes: 200 × 200 µm for SDR and 4 × 250 µm for the bipolar amplification effect (BAE) and spin dependent charge pumping (SDCP) measurements.

All EDMR measurements were made with the lab-built wafer level spectrometer. This spectrometer was built by modifying a commercially available wafer probing station. The first addition was an annular neodymium permanent magnet (B_0) hung above the wafer chuck (Fig. 2(a)), which provided the necessary magnetic field (~3300 Gauss) for X-band (~9 GHz) measurements. Second, a custom non-resonant microwave probe, shown in Figure 2(b) [1], was added to the wafer probing station. This probe introduced the oscillating (B_1) magnetic field necessary for EDMR. The microwave probe, first developed by Campbell et al. for use in scanned probe-like EPR measurements, was modified to handle more power and to generate larger B_1. The purpose of using this probe instead of the conventional microwave resonator, is to enable frequency swept measurements. Sweeping the frequency has many experimental advantages [2]-[3], and has the promise of improved sensitivity at increased microwave powers. The spectrometer is completed with the addition of four conventional electrical wafer probes. These probes are used to bias devices and measure the spin-dependent current.

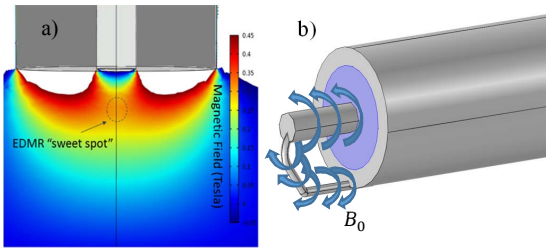


Fig. 2.(a) A finite element simulation of the neodymium permanent magnet. The illustration shows the EDMR "sweet spot" where the magnetic field is uniform over 200 µm. (b) illustrates the ~9GHz oscillating magnetic field (B_0) necessary for X-band resonance.

III. Results and Discussion

The following section presents experimental traces which demonstrate the capabilities of the wafer-level EDMR spectrometer. The wafer-level spectrometer is capable of all of the conventional biasing schemes typically used in conventional EDMR, however, these measurements are performed on a fully processed, unaltered wafer.

The transistor is biased in a variety of schemes to allow for the measurement of spin-dependent current participating in the BAE, SDCP and SDR approaches. EDMR with each of these biasing schemes was performed via frequency modulated frequency sweep.

A. Spin Dependent Recombination (SDR)

Figure 3(a) illustrates *conventional* field-swept magnetic field resonator-based EDMR of the SDR current in the source-body junction of the n-channel 4H-SiC MOSFET. The Si vacancy [7]-[8] participates quite strongly in this mode of detection resulting in a large EDMR response (S/N ≈ 20/1). This single sweep measurement was acquired in 160 seconds with a field modulation amplitude of 5 G at 1 kHz. Sample preparation for this measurement involved device identification, SiC wafer dicing, silver-paint mounting, and wire-bonding to the device contact pads. This preparation is required for every single device investigation. In terms of experimental ease of use, the wafer-level EDMR approach presents huge experimental benefits by eliminating sample preparation. While the reliability and performance limiting nature of the Si vacancy defect shown in Figure 3(a) certainly merits further study, we have instead chosen to use this defect system as a practical demonstration of the utility of the wafer-level EDMR measurement approach.

Figure 3(b) illustrates the X-band frequency-swept FM detected EDMR response of the source-substrate junction recombination current of the same SiC MOSFET illustrated in Figure 3 (a). We again observe the same Si-vacancy defect center (g = 2.003). In this measurement, the frequency sweep rate was 100 GHz/s, the FM depth was 16 MHz, and the FM rate was 10 kHz. The total acquisition time was the same (160 s) as in the conventional measurement (Fig. 3(a)). For this measurement, the input microwave power was 1 mW and the microwave probe was touching the surface of

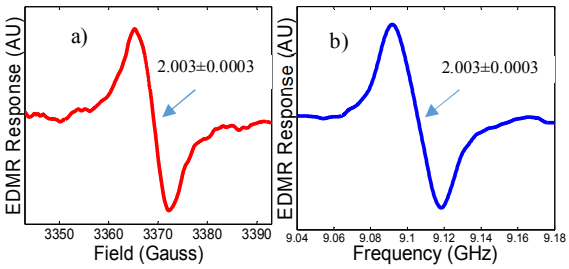


Fig. 3. (a) Conventional field swept EDMR of the 4H-SiC source-substrate. (b) FM frequency swept EDMR of the same junction. -2.5 V was applied to the source contact for both measurements.

the wafer.

A comparison of the conventional field-swept and wafer-level frequency-swept measurements can be made by realizing that 1 G is approximately equal to 2.8 MHz. Comparing the frequency swept measurement with the magnetic field swept measurement shows similar, if not slightly improved, signal to noise ratio for comparable scan settings.

B. Bipolar Amplification Effect (BAE) and Spin Dependent Charge Pumping (SDCP) Techniques

The SDR current demonstrations above provide information about deep level defects in the depletion region

Anders, Mark; Campbell, Jason; Cheung, Kin; Lenahan, Patrick; McCrory, Duane; Ryan, Jason; Shrestha, Pragya.
"Wafer Level EDMR: Magnetic Resonance in a Probing Station."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

of source to channel junctions. Variations on this measurement's biasing scheme can shift the focus from the junction to the channel region [4] and even the interface [5], [11]. These new approaches, are BAE and SDCP respectively. In the BAE biasing scheme, the transistor is biased in the sub-threshold regime such that the source-drain transport current is subject to many defect scattering events [4]. Similar to recombination in junctions, these scattering events can be rendered spin dependent at the resonance condition. Thus, the field effect transistor is biased such that it behaves like a bipolar transistor with the channel region substituting for the base region [4]. EDMR measurements in this biasing scheme provide information about the channel defects which influence transport current. As the Si-vacancy defect in SiC devices is known to dominate nearly all defect mediated currents in these devices [7]-[8], it serves as a good test for this wafer-level non-resonant EDMR measurement. Figure 4 illustrates the frequency-swept FM-modulated BAE

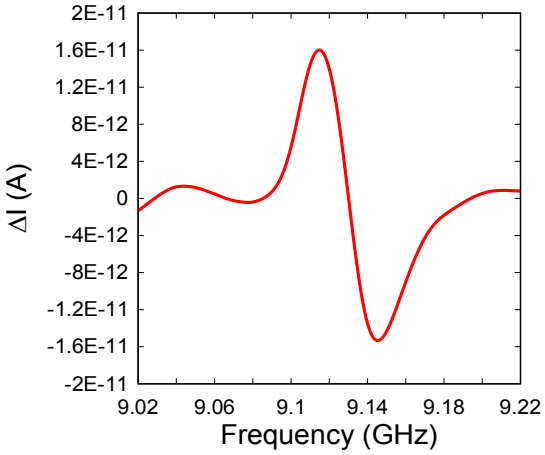


Fig. 4. EDMR of the same 4H-SiC MOSFET utilizing the BAE biasing scheme. $V_G - V_{th} = 1\,V$ and $-2.9\,V$ applied to the source of the MOSFET.

measurement on the narrow (250 µm × 4µm) SiC device.

In this measurement, the gate was biased at $V_G - V_{th} = 1\,V$, with $-2.9\,V$ on the source electrode, while the drain and substrate electrodes where grounded. This signal was acquired in 320 s with S/N ratio of ≈100. BAE is an extremely useful biasing technique that has greatly increased the signal to noise of our response, identifying the Si-vacancy defect that severely limited mobility in early SiC devices.

The previous spin-dependent current measurements, which were shown in Figures 3 and 4, were acquired in a steady state biasing scheme. The biases were chosen in order to have an approximately equal numbers of electrons and holes present for recombination. In order to increase the recombination events per unit time, we can utilize a different biasing scheme, SDCP. This active biasing scheme utilizes a square wave (50% duty cycle) that is applied to the gate electrode. The square wave cycles the device between accumulation and inversion and alternately fills the SiC/SiO_2 interface states with holes and electrons. This active biasing scheme can probe a larger energy window of defect sites relative to BAE, increasing the likelihood of identifying performance related defects within the MOSFET. Figure 5 shows SDCP detection of the Si-vacancy defect in the narrow (250 µm × 4 µm) SiC device.

For this measurement, the gate voltage waveform oscillated between +16 V and -16 V at a frequency of 200 kHz with rise/fall times of 1 µs. Detection utilized FM modulation at 10 kHz with a modulation amplitude of 8 MHz. This signal was acquired in 320 s with S/N ratio of ~10. SDCP is an extremely useful biasing scheme for probing nearly the entire energy gap of the semiconductor, while looking at defects located primarily at the interface.

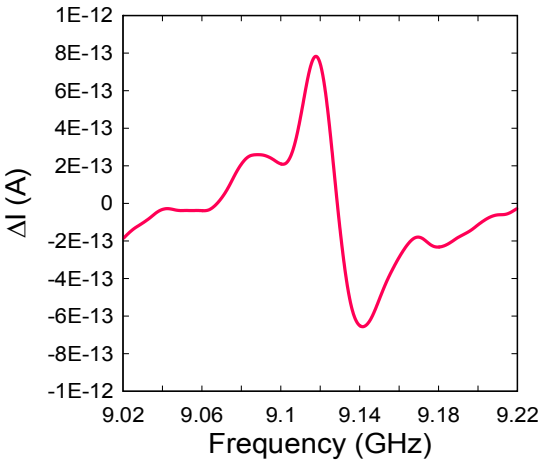


Fig. 5. SDCP of the same 4H-SiC MOSFET with ±16V square wave applied to the gate with a rise/fall time of 1µs.

This technique is also impactful due to the ability to compare the size of the SDCP response to the approximate defect counts at the interface via conventional charge pumping.

IV. Conclusion

As semiconductor technology continues to scale and new materials systems are utilized, reliability measurements need to keep pace with the advancing technology. We have demonstrated the feasibility of the wafer level EDMR spectrometer that is merged with a conventional wafer probing station. The spectrometer has demonstrated comparable, if not enhanced sensitivity in relation to conventional EDMR spectrometers. The wafer level spectrometer is capable of performing all of the conventional EDMR biasing schemes for MOSFETs. This spectrometer eliminates the need for sample preparation, performing all measurements on fully processed wafers.

The union of an EDMR spectrometer with a semiconductor wafer probing station has proven feasible and

Anders, Mark; Campbell, Jason; Cheung, Kin; Lenahan, Patrick; McCrory, Duane; Ryan, Jason; Shrestha, Pragya.
"Wafer Level EDMR: Magnetic Resonance in a Probing Station."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

promising. We believe this technique could help make EDMR a prominent technique for semiconductor reliability measurements.

ACKNOWLEDGMENT

Funding for this work was provided through the National Institute of Standards and Technology (NIST). Travel grants were provided by the EPR Technology Transfer Experience (ETTE) via SharedEPR, an NSF funded organization.

REFERENCES

[1] J. P. Campbell, J. T. Ryan, P. R. Shrestha, Z. Liu, C. Vaz, J. H. Kim, V. Georgiou, and K. P. Cheung, "Electron Spin Resonance Scanning Probe Spectroscopy for Ultrasensitive Biochemical Studies" *Anal. Chem.*, vol. 87, no. 9, pp. 4910–4916, 2015.

[2] J. W. Stoner, D. Szymanski, S. S. Eaton, R. W. Quine, G. A. Rinard, and G. R. Eaton, "Direct-detected rapid-scan EPR at 250 MHz," *J. Magn. Reson.*, vol. 170, no. 1, pp. 127–135, 2004.M.

[3] M. Tseitlin, A. Dhami, R. W. Quine, G. A. Rinard, S. S. Eaton, and G. R. Eaton, "Electron spin T-2 of a nitroxyl radical at 250 MHz measured by rapid-scan EPR," *Appl. Magn. Reson.*, vol. 30, no. 3–4, pp. 651–656, 2006.

[4] T. Aichinger and P. M. Lenahan, "Giant amplification of spin dependent recombination at heterojunctions through a gate controlled bipolar effect," *Appl. Phys. Lett.*, vol. 101, no. 8, 2012.

[5] B. C. Bittel, P. M. Lenahan, J. T. Ryan, J. Fronheiser, and A. J. Lelis, "Spin dependent charge pumping in SiC metal-oxide-semiconductor field-effect-transistors," *Appl. Phys. Lett.*, vol. 99, 2011.

[6] D. J. Lepine, "Spin-dependent transport on silicon surface," *Phys. Rev. B*, vol. 6, no. 2, pp. 436–441, 1972.

[7] C. J. Cochrane, P. M. Lenahan, and A. J. Lelis, "Deep level defects which limit current gain in 4H SiC bipolar junction transistors," *Appl. Phys. Lett.*, vol. 90, no. 12, pp. 2–5, 2007.

[8] D.J. Meyer, N.A. Bohna, and P.M. Lenahan, "Structure of 6H silicon carbide/silicon dioxide interface trapping defects" *Applied Physics Letters*, vol. 84, no. 17, pp. 3406-3408; 2004

[9] M. A. Anders, P. M. Lenahan, C. J. Cochrane, and A. J. Lelis, "Relationship Between the 4H-SiC/SiO_2 Interface Structure and Electronic Properties Explored by Electrically Detected Magnetic Resonance" *IEEE Trans. Electron Devices*, vol. 62, no. 2, pp. 301–308, 2015.

[10] M. J. Mutch, P. M. Lenahan, and S. W. King, "Defect chemistry and electronic transport in low-κ dielectrics studied with electrically detected magnetic resonance," *J. Appl. Phys.*, vol. 119, 2016.

[11] M.A. Anders, P.M. Lenahan, and A.J. Lelis, "A new electrically detected magnetic resonance approach for the study of metal-oxide-semiconductor field-effect transistor interfaces: multi resonance frequency spin dependent charge pumping and spin dependent recombination - applied to the 4H-SiC/SiO 2 interface" unpublished.

[12]] F. Klotz, H. Huebl, D. Heiss, K. Klein, J. J. Finley, and M. S. Brandt, "Coplanar stripline antenna design for optically detected magnetic resonance on semiconductor quantum dots" *Rev. Sci. Instrum.*, vol. 82, no. 7, 2011.

Anders, Mark; Campbell, Jason; Cheung, Kin; Lenahan, Patrick; McCrory, Duane; Ryan, Jason; Shrestha, Pragya.
"Wafer Level EDMR: Magnetic Resonance in a Probing Station."
Paper presented at 2017 IEEE International Integrated Reliability Workshop, South Lake Tahoe, CA, United States. October 8, 2017 - October 12, 2017.

RESONANT PRESSURE SENSING USING A MICROMECHANICAL CANTILEVER ACTUATED BY FRINGING ELECTROSTATIC FIELDS

Naftaly Krakover[1], B. Robert Ilic[2], and Slava Krylov[1]

[1]The School of Mechanical Engineering, Faculty of Engineering, Tel Aviv University, Ramat Aviv 69978, Tel Aviv, ISRAEL

[2]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA

ABSTRACT

We demonstrate a pressure-sensing approach based on the resonant operation of a single-crystal Si cantilever positioned near a flexible, pressurized membrane. The membrane deflection perturbs the electrostatic force acting on the cantilever and consequently alters the beam's resonant frequency. Sensitivity was enhanced by tailoring the actuating force nonlinearities through fringing electrostatic fields. With our coupled micromechanical system, we achieved frequency sensitivity to pressure and displacement of ≈ 30 Hz/kPa and ≈ 4 Hz/nm, respectively. Our results indicate that the suggested approach may have applications not only for pressure measurements, but also in a broad range of microelectromechanical resonant inertial, force, mass and bio sensors.

INTRODUCTION

Resonant micromechanical sensors are based on monitoring the spectral characteristics of a structure rather than static structural displacements [1]. In resonant pressure sensors, the common interrogation technique is based on a vibrating wire anchored, at its ends, to a flexible pressure-driven membrane. Deflection of the membrane results in stretching of the wire and modulation of its natural frequency [2]. In these devices, either complex force amplification mechanisms [2] or extreme dimensional downscaling [3], which may complicate fabrication, are necessary to overcome the high tensile stiffness of the vibrating wire and to improve sensitivity. Moreover, frequency of doubly-clamped wires is affected by residual and thermal stresses.

Micro- and nano-scale devices incorporating cantilevers as the sensing elements have numerous applications in diverse fields ranging from engineering to life and physical sciences. These sensors are used for the detection of small masses, biomolecular binding events, and non-contact topographic and localized charge imaging using atomic force microscopy [4]. The key advantage of cantilever-based devices are their lower (when compared to doubly-clamped beams and wires) stiffness and reduced sensitivity to temperature and residual stress.

Despite their widespread use, vibrating cantilevers have not yet been implemented in resonant pressure sensors. A cantilever sensor exploiting the effect of ambient pressure on the frequency of a resonator was reported in [5]. Cantilevers were placed in an evacuated chamber and resonant frequencies were measured at values of pressure varying between 0.1 Pa to 10^5 Pa. The use of this kind of device for open-air pressure measurements could be challenging since the cantilever response is also influenced by other factors such as temperature, humidity and contamination. In the present work, we introduce a membrane-based pressure sensor, which incorporates an oscillating cantilever as the sensing element. To ensure controlled environmental conditions, the cantilever can be positioned in a sealed cavity under the membrane, whereas the other side of the membrane is exposed to ambient conditions.

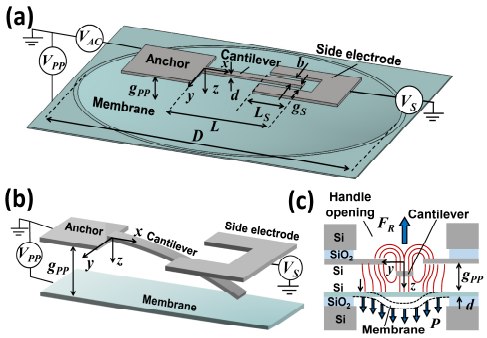


Figure 1: Schematic illustration of the device. (a) The cantilever is positioned above a pressure-driven, deformable membrane. The handle wafer that the cantilever is attached to is not shown. (b) Deformed cantilever is actuated by the parallel plate PP and side electrodes. (c) Device cross section showing the cantilever, the side electrode and the membrane. The restoring force F_R appears due the asymmetry of the fringing fields (thin red lines). Dashed line shows the deformed membrane.

Significant research efforts have been devoted to design and measurement method improvements, as reviewed in [4]. The most commonly explored sensing scenarios primarily focus on the linear operating regimes of the vibrating structure. One of the emerging strategies for sensitivity enhancement is based on device operation in a nonlinear regime. In these structures, especially when actuation takes place near instability points, small variations in the detection parameter may lead to a large response changes, therefore improving sensitivity [6-9]. Electrostatically actuated devices manifest the so-called pull-in instability associated with the nonlinearity of the actuating forces. Operation near the pull-in point, where the effective device stiffness is minimal due to the electrostatic softening, increases frequency sensitivity [7]. However, operation near the instability has drawbacks

Ilic, Bojan; Krakover, Naftaly; Krylov, Slava.
"Resonant Pressure Sensing Using a Micromechanical Cantilever Actuated by Fringing Electrostatic Fields."
Paper presented at The 31st IEEE International Conference on Micro Electro Mechanical Systems, MEMS 2018, Belfast, United Kingdom.
January 21, 2018 - January 25, 2018.

including a possibility of device collapse into the electrode and irreversible damage.

An alternative approach for sensitivity enhancement of cantilever-type architectures operating near critical points is based on the actuation by fringing electrostatic fields [10]. This kind of actuation introduces an additional nonlinearity that can lead to an inflection point in the voltage-deflection dependence, where the cantilever is on the verge of bistability. It was shown theoretically [10] that device operation near the critical point may lead to sensitivity enhancement by more than an order of magnitude.

In this work, we present a pressure sensor based on a cantilever-near-a-membrane type resonant device. To enhance sensitivity, the cantilever is actuated near an inflection point by the electrostatic fringing fields and the parallel-plate capacitive forces. The suggested actuation mechanism can be used as a general approach for other displacement sensing devices.

DEVICE ARHITECTURE AND OPERATIONAL PRINCIPLE

The device die, containing a cantilever of length L, width b, and thickness d, is attached to the flexible membrane of diameter D. The initial distance between the cantilever and the membrane, defining the gap within the parallel plate (PP) capacitive actuator, is g_{PP} (Fig. 1). The planar side (S) electrode of length L_S, located at a distance g_S from the cantilever (Fig. 1a), is a source of the restoring electrostatic force F_R associated with the fringing fields [10] (Fig 1c). A pressure P causes the membrane to deflect. This leads to an increase of the gap g_{PP} between the membrane and the cantilever.

In general, the electrostatic force acting on the beam is affected by the interaction between the S and the PP electrodes [10] and is more complicated than just a superposition of the forces. To model the effects, we first considered a case without a PP electrode, when the interaction is solely between the cantilever and the side electrode. The force f_s (per unit length of the beam) provided by the side electrode can be approximated by the expression [11]

$$f_s(x,t) = -\frac{\alpha\sigma(w/d)V_s^2}{1+\sigma|w/d|^\gamma} \qquad (L_S \le x \le L) \qquad (1)$$

where $w(x,t)$ is the deflection of the beam, x and t are the coordinate along the beam and time, respectively, V_S is the voltage on the side electrode and α, σ, γ are fitting parameters. Due to symmetry, the resultant electrostatic force is zero in the initial configuration. In the deflected state, the distributed electrostatic force, arising from asymmetries of the fringing fields, acts in a direction opposite to the beam's deflection and effectively serves as a restoring force [11]. In contrast, the nonlinear force provided by the parallel plate electrode, which can be approximated by the simplest PP capacitor formula,

$$f_{PP}(x,t) = \frac{\varepsilon_0 b V_{PP}^2}{2(g_{PP}-w)^2} \qquad (2)$$

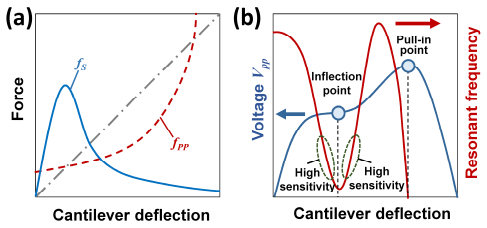


Figure 2: (a) Schematics of the mechanical (dash-dot line), electrostatic restoring forces f_S (solid line) and the electrostatic force f_{PP} (dashed line) as a function of the cantilever's deflection. (b) Resulting equilibrium curve (blue) and the resonant frequency of the beam (red) as a function of the beam's deflection.

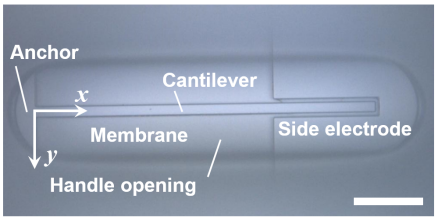


Figure 3: Optical micrograph of the integrated device. The image of the cantilever and of the side electrode is acquired through the opening in the handle wafer. Scale bar size is ≈ 100 μm.

is of a divergent nature and consequently pulls the beam further away from its initial configuration. Here, ε_0 is the dielectric permittivity and V_{PP} is the voltage applied to the PP electrode. The force f_{PP} is affected by the gap between the cantilever and the membrane, and therefore by the membrane deflection. Figure 2a shows schematically the qualitative dependence of the elastic, side force f_s and PP forces f_{PP} as a function of the beam's deflection (at $x = L$). An equilibrium curve for a specific value of V_S, is shown in Fig 2b. The value of V_S can be chosen in such a way that the force-deflection curve contains an inflection point. In the vicinity of this point, both the effective stiffness and the frequency decrease, whereas the slope of the frequency-deflection curve (red line in Fig. 2b) becomes steeper. Changes in pressure P modify the gap g_{pp} and affect the natural frequency of the beam. By choosing appropriate V_S and V_{PP} such that the beam is positioned near the inflection point, frequency sensitivity of the beam to the membrane deflection can be significantly improved. We emphasize that Eqs. (1) and (2) are used here only to show a qualitative dependence between the frequency and the membrane deflection and to clarify the sensor's operational principle. Numerical approaches should be used for more accurate evaluation of this force [10,11].

Ilic, Bojan; Krakover, Naftaly; Krylov, Slava.
"Resonant Pressure Sensing Using a Micromechanical Cantilever Actuated by Fringing Electrostatic Fields."
Paper presented at The 31st IEEE International Conference on Micro Electro Mechanical Systems, MEMS 2018, Belfast, United Kingdom.
January 21, 2018 - January 25, 2018.

EXPERIMENT

Using deep reactive ion etching (DRIE), cantilevers, side electrodes and the membrane were fabricated from a $\approx 5\ \mu$m thick, single crystal, silicon device layer using a silicon-on-insulator (SOI) wafer with a $\approx 2\ \mu$m thick buried silicon dioxide layer. DRIE was also used to etch a cavity within the handle wafer to form the membrane and an opening in the handle under the cantilever to allow for large amplitude vibrations of the beam. Using a polymer spacer of thickness g_{PP}, the cantilever die was flipped upside down and attached to the membrane, in such a way that the device is facing the membrane, Fig 1c. Figure 3 shows an optical micrograph of the fabricated device with the dimensions $L \approx 1000\ \mu$m, $b \approx 16\ \mu$m, $d \approx 5\ \mu$m, $g_S \approx 5\ \mu$m, $L_S \approx 5\ \mu$m, $g_{PP} \approx 10\ \mu$m, and $D \approx 2000\ \mu$m. The experimental setup is presented in Fig. 4.

The assembly containing both chips, each with an approximate extent of 1 cm × 1 cm, were mounted onto a custom built printed circuit board (PCB). The beam and the side electrode were wire-bonded to the PCB contact pads. The double chip and the PCB stack assembly were placed onto a holder in such a manner that the membrane was positioned above a sink hole. The hole was connected by a tube to a pump applying a suction pressure in the range of $P \approx 12$ kPa to $P \approx 82$ kPa. The cantilever side of the membrane was at the atmospheric pressure.

The sinusoidal, zero offset, voltage signal with an amplitude of $V_{AC} \approx 2$ V, provided by a network analyzer, was supplied to the cantilever. The frequency was swept between ≈ 29 kHz and ≈ 34 kHz. In addition, using a separate power supply, steady-state voltages V_{PP} and V_S were applied to the membrane and the side electrodes, respectively. Using a single-beam laser Doppler vibrometer (LDV) operated in a velocity acquisition mode, the out-of-plane response of the cantilever was measured. The positioning of the LDV laser spot was monitored by a camera mounted on the microscope. The output of the LDV was fed back into the network analyzer. In parallel to the spectral analysis of the output signal, the velocity time history of the LDV output was monitored with an oscilloscope.

An example of frequency response for $V_S \approx 50$ V,

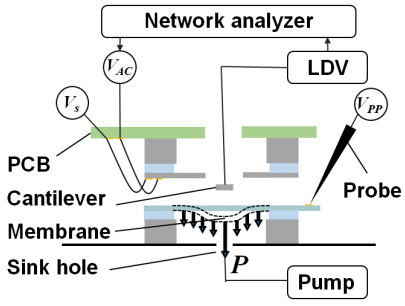


Fig. 4. Schematic illustration of the experimental setup. The device stack contains two SOI chips positioned on a PCB board, placed in such a way that the membrane is positioned above the sink hole. Cantilever vibrations were measured using the LDV through an opening in the handle of the SOI wafer and the PCB. Spectral response was measured using a network analyzer.

$V_{pp} \approx 30$ V and a varying pressure P is shown in Fig. 5. The measured spectra show a resonant frequency increase with increasing pressure. The experiments were carried out under various combinations of V_S and V_{pp}. Fig. 6a shows the resonant frequency f dependence on P for $V_{pp} \approx 30$ V and $V_{pp} \approx 70$ V. The slope df/dP of the curves represents the frequency sensitivity to pressure, as shown in the Fig. 6a inset. We found that higher sensitivity is reached for lower pressure values. This corresponds to a smaller gap g_{PP} and consequently larger PP capacitive force. In our experiments, we measured a sensitivity of

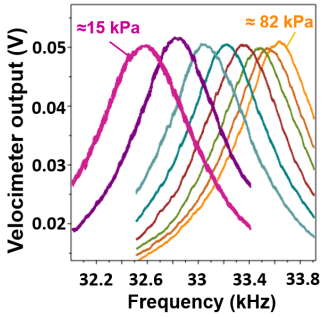


Fig 5. Measured resonant curves of the cantilever beam at a varying P for $V_S \approx 50$ V and $V_{PP} \approx 30$ V.

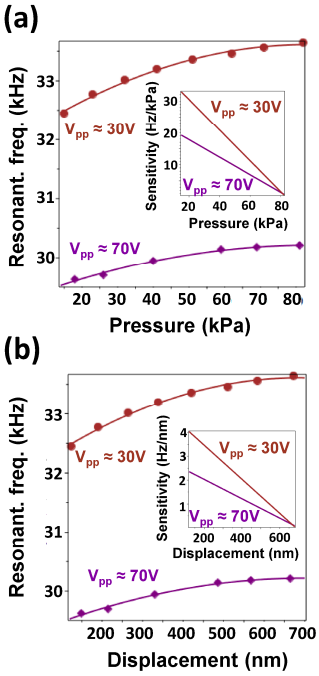


Fig. 6. (a) Resonant frequency as a function of the applied pressure at two values of V_{pp}. The slope of the response curve represents frequency sensitivity to pressure (inset). The one standard deviation frequency uncertainty based on a Lorentzian fit is smaller than the data marker size. (b) Resonant frequency as a function of membrane displacement and the corresponding frequency sensitivities to displacement (inset).

Ilic, Bojan; Krakover, Naftaly; Krylov, Slava.
"Resonant Pressure Sensing Using a Micromechanical Cantilever Actuated by Fringing Electrostatic Fields."
Paper presented at The 31st IEEE International Conference on Micro Electro Mechanical Systems, MEMS 2018, Belfast, United Kingdom.
January 21, 2018 - January 25, 2018.

≈ 30 Hz/kPa of suction pressure under the membrane. This result is comparable to state-of-the-art values reported in the literature [2,12].

The suggested approach can be considered as a particular case of a generic displacement sensing method. In order to estimate the measured frequency sensitivity to the PP electrode displacement, membrane midpoint deflection w_0 was calculated using the expression [13]

$$w_0 = \frac{3PD^4(1-\nu^2)}{256Ed^3\left(1+\frac{0.488w_0^2}{d^2}\right)} \quad (3)$$

Here, E and ν are Young's modulus and the Poisson's ratio of the membrane material, respectively. By calculating the deflection for each value of P that the frequency was measured for, frequency sensitivity (scale factor) curves were obtained (Fig 6b). The highest measured frequency sensitivity-to-displacement was ≈ 4 Hz/nm.

CONCLUSION

We demonstrate a cantilever-membrane resonant pressure sensing approach. Deflection of a membrane, resulting from pressure differences on its two sides, changes the nonlinear electrostatic force acting on the cantilever and consequently alters the cantilever's natural frequency. The use of the fringing electrostatic field actuation, in addition to the force provided by the parallel-plate actuator, allowed for tailoring of the effective nonlinearity of the system that in turn enhanced sensitivity. Along with pressure sensing, the suggested design can be used as a general approach to frequency-based displacement sensing. Our results indicate that the integrated sensing platform based on a coupled mechanical system may be applicable for a wide range of displacement sensing applications, ranging from pressure sensors and accelerometers to flow, tactile, inertial and precision force sensors.

ACKNOWLEDGEMENTS

The devices were fabricated and tested at the Center for Nanoscale Science & Technology (CNST) at the National Institute of Standards and Technology (NIST) and at the Microsystems Design and Characterization Laboratory at Tel Aviv University.

REFERENCES

[1] R. Abdolvand, B. Bahreyni, J. Lee, and F. Nabki, "Micromachined Resonators: A Review," *Micromachines*, vol. 7, no. 9, p. 160, Sep. 2016.

[2] X. Du, L. Wang, A. Li, L. Wang, and D. Sun, "High Accuracy Resonant Pressure Sensor With Balanced-Mass DETF Resonator and Twinborn Diaphragms," *J. Microelectromech. Syst.*, vol. 26, no. 1, pp. 235–245, Feb. 2017.

[3] M. Messina, J. Njuguna, V. Dariol, C. Pace, and G. Angeletti, "Design and Simulation of a Novel Biomechanic Piezoresistive Sensor With Silicon Nanowires," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 3, pp. 1201–1210, Jun. 2013.

[4] A. Boisen, S. Dohn, S. S. Keller, S. Schmid, and M. Tenje, "Cantilever-like micromechanical sensors," *Reports Prog. Phys.*, vol. 74, no. 3, p. 36101, 2011.

[5] S. Bianco, M. Cocuzza, S. Ferrero, E. Giuri, G. Piacenza, C. F. Pirri, A. Ricci, L. Scaltrito, D. Bich, A. Merialdo, P. Schina, and R. Correale, "Silicon resonant microcantilevers for absolute pressure measurement," *J. Vac. Sci. Technol. B.*, vol. 24, no. 4, p. 1803, 2006.

[6] N. Krakover, B. R. Ilic, and S. Krylov, "Displacement sensing based on resonant frequency monitoring of electrostatically actuated curved micro beams," *J. Micromech. Microeng.*, vol. 26, no. 11, p. 115006, Nov. 2016.

[7] C. Comi, A. Corigliano, A. Ghisi, and S. Zerbini, "A resonant micro accelerometer based on electrostatic stiffness variation," *Meccanica*, vol. 48, no. 8, pp. 1893–1900, 2013.

[8] A. Ramini, M. I. Younis, and Q. T. Su, "A low-g electrostatically actuated resonant switch," *Smart Mater. Struct.*, vol. 22, no. 2, p. 25006, Feb. 2013.

[9] M. E. Khater, M. Al-Ghamdi, S. Park, K. M. E. Stewart, E. M. Abdel-Rahman, A. Penlidis, A. H. Nayfeh, A. K. S. Abdel-Aziz, and M. Basha, "Binary MEMS gas sensors," *J. Micromec. Microeng.*, vol. 24, no. 6, p. 65007, Jun. 2014.

[10] N. Krakover and S. Krylov, "Bistable Cantilevers Actuated by Fringing Electrostatic Fields," *J. Vib. Acoust.*, vol. 139, no. 4, p. 40908, May 2017.

[11] Y. Linzon, B. Ilic, S. Lulinsky, and S. Krylov, "Efficient parametric excitation of silicon-on-insulator microcantilever beams by fringing electrostatic fields," *J. Appl. Phys.*, vol. 113, no. 16, p. 163508, 2013.

[12] Z. Tang, S. Fan, and C. Cai, "A silicon micromachined resonant pressure sensor," *J. Phys. Conf. Ser.*, vol. 188, no. 1, p. 12042, Sep. 2009.

[13] S. P. Timoshenko and S. Woinowsky-Krieger, *Theory of plates and shells*. McGraw-hill, 1959.

Ilic, Bojan; Krakover, Naftaly; Krylov, Slava.
"Resonant Pressure Sensing Using a Micromechanical Cantilever Actuated by Fringing Electrostatic Fields."
Paper presented at The 31st IEEE International Conference on Micro Electro Mechanical Systems, MEMS 2018, Belfast, United Kingdom.
January 21, 2018 - January 25, 2018.

Errors in Rate-of-Rise Gas Flow Measurements from Flow Work

John D. Wright, Aaron N. Johnson, Michael R. Moldover, and Gina M. Kline
National Institute of Standards and Technology (NIST), 100 Bureau Drive, Gaithersburg, MD, USA 20899
Corresponding Author: john.wright@nist.gov

Abstract
The rate-of-rise (RoR) method determines flow by measuring the time rate of change of the amount of gas in a collection tank of known volume as it is filled via a flow meter under test. The mass of gas is calculated from time-stamped pressure and temperature data and accurate RoR measurements require reliable gas pressure and temperature values while the collection tank is filling with gas. We present a thermodynamic model and experimental measurements of gas temperature errors that are a function of a dimensionless ratio related to: {the heat transfer from the gas to its surroundings} / {the heat generated by flow work}. The uncertainty of RoR flow measurements made using the NIST 34 L collection tank is <0.12 % for flows between 1 sccm* and 200 sccm. At lower and higher flows, the RoR uncertainty rises to approximately 1 % due to leaks and temperature errors induced by flow work.

1. Introduction

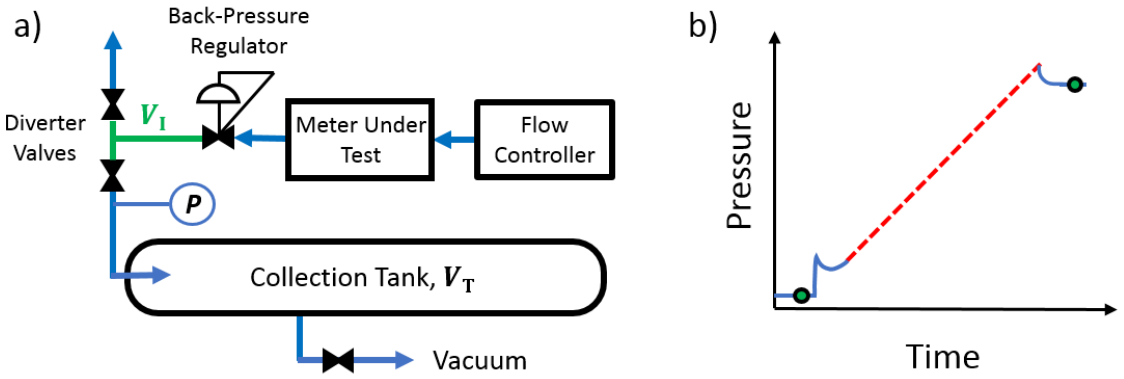


Figure 1. a) Components of NIST's *PVTt* and Rate-of-rise (RoR) flow standard. b) Time-dependent pressure measured during a calibration by both methods. For the RoR calibration, the mass flow $\dot{m}_{\mathrm{RoR}}$ is calculated from the slope of the red dashed line; the initial and final transients (blue curves) are ignored. For a *PVTt* calibration, the initial and final masses are calculated from the pressure data before and after the transients (green circles). The mass difference is divided by the total filling time.

* sccm = standard cubic centimeters per minute (cm^3/min) with reference conditions of 101.325 kPa and 0° C and slm = standard liters per minute (L/min) at the same reference conditions.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

The rate-of-rise (RoR) method calculates the rate of change of mass during the tank filling process via the equation:

$$\dot{m}_{\mathrm{RoR}} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\left(V_{\mathrm{T}} + V_{\mathrm{I}}\right)\rho(P,T)\right] = \frac{N\sum_{j=1}^{N} t_j m_j - \sum_{j=1}^{N} t_j \sum_{j=1}^{N} m_j}{N\sum_{j=1}^{N} t_j^2 - \left(\sum_{j=1}^{N} t_j\right)^2}\ , \qquad (1)$$

where V_{T} is the collection tank volume, V_{I} is the inventory volume between a back-pressure regulator and the diverter valves, P and T are the pressure and temperature of the collected gas, and t is time. The real gas density $\rho(P,T)$ can be calculated from the NIST properties database REFPROP [1]. The latter part of Equation 1 is the slope of the mass-versus-time record, as determined using a first order least squares regression [2] with N mass values evenly spaced in time where $m_j = \left(V_{\mathrm{T}} + V_{\mathrm{I}}\right)\rho\left(P_j, T_j\right)$. The RoR method is a dynamic volumetric method because the pressure and temperature measurements are made while their values are changing, not at steady state. RoR can be applied to a collection tank that is used as a sink (as described above) or as a source. A similar apparatus can be used to make Pressure-Volume-Temperature-time (*PVTt*) flow measurements, as explained in Section 2.

To calibrate a flow meter, data from the meter under test are collected and averaged over the same time interval that the RoR or *PVTt* measurements are made. Valves allow the collection tank to be evacuated by a vacuum pump and then filled with a flow through the meter under test (Figure 1). Note that for RoR measurements, the collection volume is $V_{\mathrm{T}} + V_{\mathrm{I}}$, where V_{I} is the inventory volume.

The RoR method has been used by the semiconductor manufacturing industry for many years to calibrate mass flow controllers using chemical vapor deposition chambers as the collection tank [3]. NIST has applied the RoR method at flows below 100 sccm using our 34 L *PVTt* collection tank. When used for *PVTt* measurements, the 34 L collection tank is normally filled from vacuum to 100 kPa, but if this were done for a flow of 1 sccm, the filling would take 24 days! The RoR method allows us to gather calibration data for small gas flows in a shorter time, so that a typical RoR calibration at one flow set point will gather 1 h of pressure and temperature data at 10 second intervals.

In this paper, we will document details of the RoR data collection and processing, explain the reasons for observed differences between RoR and *PVTt* flow measurements made at NIST, and give an uncertainty analysis for the RoR method. Our goal is to provide guidance on how to design and use a RoR standard to obtain low uncertainty measurements of small gas flows.

2. NIST Implementation of *PVTt* Standards

The 34 L *PVTt* gas flow standards measure the time to fill the collection tank from 0.020 kPa to a "full" condition, usually 100 kPa. The operation and uncertainty of the *PVTt* standards is documented in detail elsewhere [4, 5] and is only briefly summarized here. Diverter valves are used to switch flow from the meter under test into the collection tank. The temperature and pressure measurements of the gas in the collection tank under steady state conditions, before and after filling, are used to calculate the gas density (green circles in Figure 1). Flow is calculated using the equation:

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$\dot{m}_{PVTt} = \frac{V_{\mathrm{T}}\left(\rho_{\mathrm{T}}^{(f)} - \rho_{\mathrm{T}}^{(i)}\right) + V_{\mathrm{I}}\left(\rho_{\mathrm{I}}^{(f)} - \rho_{\mathrm{I}}^{(i)}\right)}{t^{(f)} - t^{(i)}} \quad , \tag{2}$$

where (i) and (f) indicate initial and final values respectively. The *PVTt* method must account for pressure and temperature changes in the inventory volume between the meter under test and the collection tank. At NIST this is done by using an inventory "mass cancellation" technique that ensures that the initial and final density in the inventory volume are equal, $\rho_{\mathrm{I}}^{(f)} = \rho_{\mathrm{I}}^{(i)}$ [5]. In contrast, the RoR method is applied during intervals where pressure and temperature in the collection tank are unaffected by the switching of the diverter valves (red dashed line in Figure 1). The RoR method has the advantage that it does not use the extra instrumentation, data acquisition, and processing required to implement mass cancellation. In addition, for a collection tank of a given size, RoR enables low flow measurements in a much shorter time than static flow measurement standards such as the *PVTt* method.

The volume of the 34 L collection tank was measured gravimetrically with an uncertainty of < 0.012 %*. The NIST *PVTt* flow standards and their uncertainty of 0.025 % have been validated by numerous inter-laboratory comparisons. Here, the *PVTt* flow measurements will be used as the reference to evaluate the RoR method.

The 34 L collection tank is submerged in a temperature controlled water bath with a set point equal to the nominal room temperature, $T_{\mathrm{H_2O}} = 296.463$ K. A proportional-integral-derivative feedback controller, recirculation pumps, and baffles provide temporal and spatial uniformity of the water bath temperature of ±0.002 K. It is difficult to accurately measure the temperature of the gas in the collection tank because of the low heat capacity of the gas relative to temperature sensors and the poor heat transfer between them. It is easier to make a low uncertainty temperature measurement of the recirculating, high-heat-capacity water bath. For *PVTt* flow measurements, we wait 10 minutes after filling to achieve thermal equilibrium between the gas and the surrounding water and measure $T_{\mathrm{H_2O}}$ to obtain the temperature of the gas in the collection tank with an uncertainty of 0.012 K. We note that the collection tank has a diameter of only 15 cm to improve heat transfer and reduce the time necessary to obtain equilibrium.

The time constant for thermal equilibrium of the gas in the full collection tank is <170 s. It was determined by filling the collection tank and recording the time necessary for the pressure to equilibrate (Figure 2). Flow work (heat of compression) elevates the temperature of the gas in the collection tank during filling [6]. After the filling stops, the pressure declines as the gas returns to thermal equilibrium with the water bath. Mass conservation dictates that the average gas density remains constant during the thermal equilibration process. The decreasing pressure correlates with the decreasing gas temperature, and the time required for pressure stabilization determines the thermal time constant.

Figure 2 illustrates equilibration following a flow of 5 slm. In this case, the temperature of the gas is approximately 2.5 K warmer than the water bath after filling. The gas temperature while the

* All uncertainties herein are *k* = 2, approximately 95 % confidence level.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

collection tank is filling is an important uncertainty component for RoR flow measurements. A thermodynamic model for the temperature difference between the gas and water T_{err} (see Section 5) reveals that T_{err} is related to a dimensionless ratio that determines the heat transfer from the gas to its surroundings relative to the heat generated by flow work.

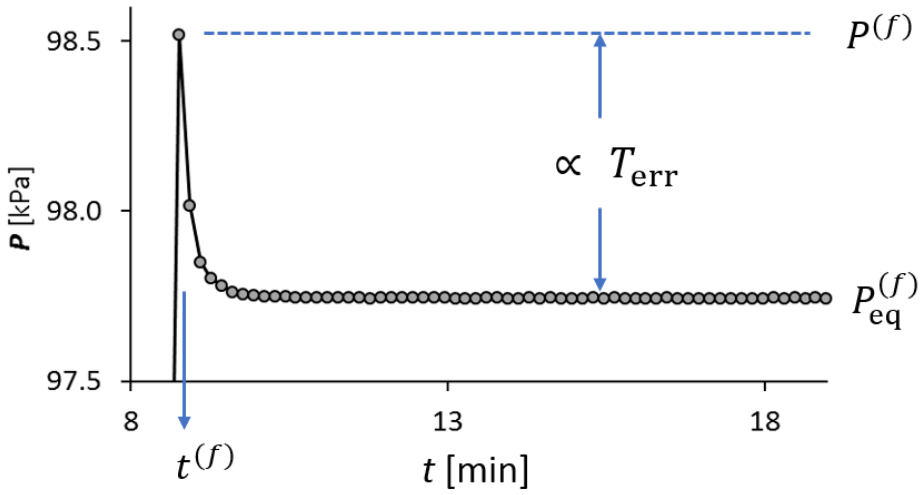


Figure 2. Pressure in the 34 L collection tank as a function of time for an air flow of 5 slm. The gas is initially warm (and therefore at higher pressure) due to flow work during filling. After filling stops at the time $t^{(f)}$, the gas temperature comes to equilibrium with the water bath with a time constant of 170 s.

3. Comparison of RoR and *PVTt* Flow Measurements

Figure 3 shows the difference between *PVTt* and RoR methods when applied to the same data sets. The RoR method is applied over a shorter time interval than the *PVTt* method: transient pressure data caused by opening the diverter valve are excluded from the RoR processing. The effects of the different time intervals have been removed by using data from the meter under test as an intermediary to normalize the data from the two techniques. The RoR flows in Figure 3 assume that the gas is in thermal equilibrium with the water bath and there is no pressure drop between the pressure sensor and the tank (discussed in Section 4). The RoR results agree with *PVTt* flows near zero, but the difference increases linearly to 0.3 % near 1000 sccm. Because of data like these, the NIST RoR system has only been used to report flows < 100 sccm. In this paper, we develop a RoR uncertainty analysis that explains the differences in Figure 3 and shows how to avoid them. For the purposes of this research, we used the 34 L tank to make RoR measurements up to 10 slm, one hundred times the normal maximum.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

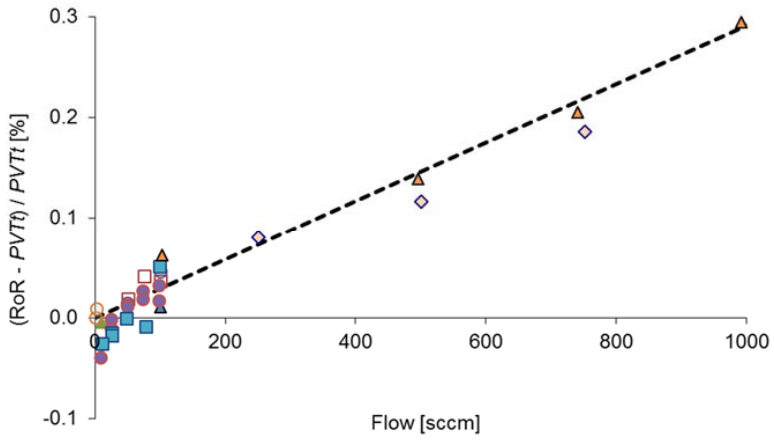


Figure 3. Fractional differences between Rate-of-rise and *PVTt* flow measurements made with the 34 L collection tank. The dashed line shows the fractional differences between the two methods predicted by temperature errors caused by flow work, T_{err} (dashed line in Figure 6).

Assuming ideal gas behavior, a stable zero offset in the pressure measurement introduces no error in RoR flow measurements. Density is proportional to pressure and, because the slope calculated in Equation 1 depends on pressure differences, a zero offset in pressure cancels. However, density varies inversely with temperature, and therefore a zero offset in temperature does not cancel. A temperature offset T_{err} causes a fractional RoR mass flow error of $T_{\mathrm{err}}\ /T$. Our experiments show that the difference between the RoR and *PVTt* flows shown in Figure 3 is predominantly caused by errors in the temperature measurement of the gas while the tank is filling. The flow work and resulting temperature errors increase with increasing flow. In Section 5 we will study the temperature errors in more detail.

4. Pressure Errors, P_{err}

Although zero offsets in pressure do not lead to RoR flow uncertainty, any pressure error that does not remain constant while the tank is filling will introduce RoR flow error. Two examples of non-constant pressure errors are: (1) gain (span) errors of the sensor calibration, and (2) non-linearity in the pressure measurements. Here non-linearity means that the measured pressure is not proportional to the actual pressure in the tank. The pressure sensors used in the NIST 34 L flow standard have excellent linearity under steady state pressure conditions; however, their pressure readings during filling depend on the flow and the pressure in the tank. The piping between the pressure sensor and the tank causes a pressure drop (P_{err}) that decreases as the tank fills.* The changing pressure drop introduces a non-linearity to the pressure measurements that cause a RoR flow error. Another possible cause of significant pressure non-linearity, addressed in Section 5, is an increase in the sensor's temperature due to flow work.

* The head correction because the pressure sensor is not located at the same elevation as the tank is another such source of non-linearity, but it is < 0.01 % for the 34 L system.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

To accurately estimate the pressure drop, we constructed a full-scale model of the inlet piping (Figure 4a). The model reproduced the lengths of pipe and elbows between the tap for the pressure sensors and the collection tank. We measured the pressure drop with a differential pressure sensor for flows between 1 slm and 10 slm and for "tank" pressures between 11 kPa and 100 kPa (Figure 4b and c). Both flow and tank pressure influence the pressure drop. The pressure errors (and resulting RoR flow errors) are larger for larger flows and for smaller tank pressures. Using the data in Figure 4b, a RoR flow measurement at 10 slm that filled the tank from 11 kPa to 100 kPa will have a non-linearity introduced by the changing pressure drops of 137 Pa – 16 Pa = 121 Pa. This pressure non-linearity will lead to a flow measurement error of 0.121 kPa / 89 kPa = 0.14 %. If data for pressures between 30 kPa and 100 kPa are used instead, the RoR flow error is reduced to 0.035 kPa / 100 kPa = 0.05 %. If the pressure drop as a function of flow and tank pressure were well characterized, it would be practical to correct the pressure measurements. Also, a better position of the pressure tap (i.e. on the tank itself instead of the inlet piping) would probably avoid significant errors due to pressure drop. We have not yet done so because the plumbing change would necessitate a new volume determination for the 34 L tank.

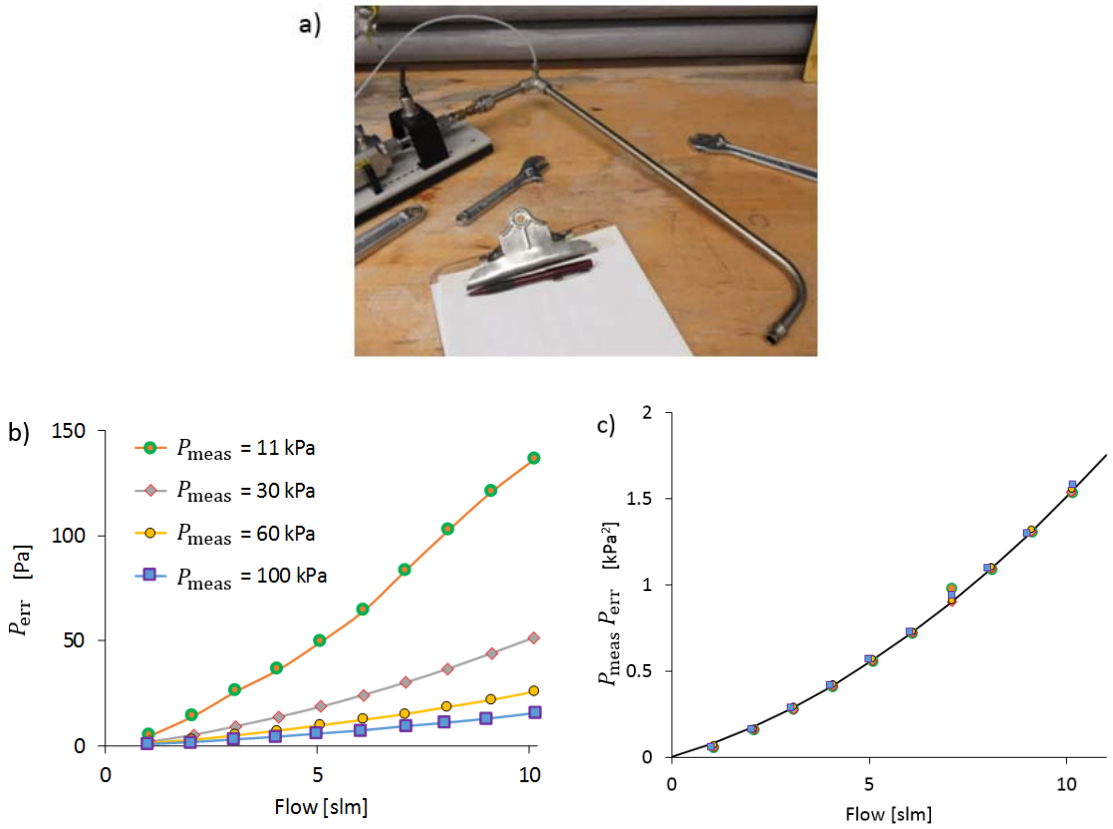


Figure 4. a) The reproduction of the 34 L tank inlet piping, b) the pressure drop *versus* flow for various tank pressures, and c) the product of the measured tank pressure and pressure drop *versus* flow.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

The function,

$$P_{\mathrm{err}} = b_1 \frac{\dot{m}}{P_{\mathrm{meas}}} + b_2 \frac{\dot{m}^2}{P_{\mathrm{meas}}} \quad , \tag{3}$$

was used to fit the experimental data in Figure 4 with residual standard deviation of 2.2 Pa. In Equation 3, $\dot{m}$ is the mass flow, P_{meas} is the measured tank pressure, and b_1 and b_2 are best-fit coefficients. The functional form of Equation 3 is expected for laminar flow in a tube with other loss sources (in this case elbows). The P_{err} measurements are based on a reproduction of the tank inlet plumbing, not the actual piping, so they are not used to make corrections to the RoR pressure measurements. Here, the P_{err} fit is used to experimentally determine temperature measurement errors and during the RoR uncertainty analysis.

5. Temperature Errors from Flow Work, T_{err}

Flow work heats the gas in the collection tank during filling and causes the gas temperature to exceed that of the water bath. The availability of *PVTt* flow measurements from the NIST 34 L system allows experimental assessment of these temperature errors. The experimental data validate our thermodynamic model of flow work for the RoR standard and guide future RoR system design and application.

The measurements of temperature errors use the collection tank as a "gas thermometer": assuming a constant mass flow entering the tank, we can use the RoR and *PVTt* data to calculate the pressure that would be present in the tank P_{eq} if the gas were in equilibrium with the surrounding water bath. Then the ratio of the measured tank pressure to P_{eq} can be used to calculate the actual average temperature of the gas during filling:

$$T = \frac{P_{\mathrm{T}}}{P_{\mathrm{eq}}} T_{\mathrm{H_2O}} = \frac{(P_{\mathrm{meas}} - P_{\mathrm{err}})}{P_{\mathrm{eq}}} T_{\mathrm{H_2O}} \quad , \tag{4}$$

where $T_{\mathrm{H_2O}}$ is the temperature of the water bath, P_{meas} is the pressure measured by the calibrated pressure transducers connected to the collection tank, P_{err} is the pressure drop calculated from Equation 3, and $P_{\mathrm{T}} = P_{\mathrm{meas}} - P_{\mathrm{err}}$ is our best estimate of the pressure in the filling tank. (The gas is assumed to be ideal throughout this analysis.)

The estimate of P_{eq} is determined so that the RoR flow matches the *PVTt* flow calculation. For a steady flow of an ideal gas, $P_{\mathrm{eq}}(t)$ varies linearly with time and is expressed by

$$P_{\mathrm{eq}}(t) = P_{\mathrm{eq}}^{(i)} + \frac{\mathrm{d}P}{\mathrm{d}t}\left(t - t^{(i)}\right) = P_{\mathrm{eq}}^{(i)} + \left[\frac{P_{\mathrm{eq}}^{(f)} - P_{\mathrm{eq}}^{(i)}}{t^{(f)} - t^{(i)}}\right]\left(t - t^{(i)}\right) \quad , \tag{5}$$

where the superscripts (i) and (f) indicate the initial (start) and final (stop) conditions of the collection. The equilibrium pressure of the full tank $P_{\mathrm{eq}}^{(f)}$ is available by waiting for the gas to return to thermal equilibrium with the surrounding water (see Figure 2 or 4).

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

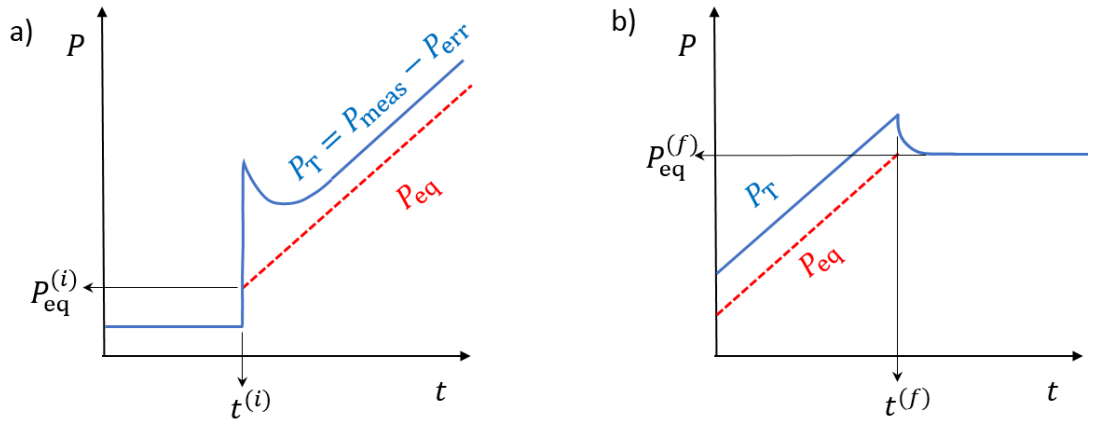


Figure 5. Plots of pressure versus time for the a) start and b) stop of the collection tank filling labelled with quantities used to experimentally determine temperature errors T_{err} caused by flow work in the RoR system.

The initial equilibrium pressure $P_{\mathrm{eq}}^{(i)}$ is not equal to the pressure in the collection tank before filling because there is a sudden in-rush of gas from the inventory volume. Hence $P_{\mathrm{eq}}^{(i)}$ includes mass from three sources:

$$P_{\mathrm{eq}}^{(i)} = \frac{\left[m_{\mathrm{T}}^{(i)} + m_{\mathrm{I}}^{(i)} + m_{\mathrm{DE}}^{(i)}\right] R_{\mathrm{u}} T_{\mathrm{H_2O}}}{M(V_{\mathrm{T}}+V_{\mathrm{I}})} \ , \tag{6}$$

where R_{u} is the universal gas constant, and M is the gas molar mass. The initial mass of gas in the collection tank $m_{\mathrm{T}}^{(i)}$ and in the inventory volume $m_{\mathrm{I}}^{(i)}$ are calculated from pressure and temperature values gathered for the *PVTt* flow measurements, before the gas is diverted to the tank, under conditions of thermal equilibrium:

$$m_{\mathrm{T}}^{(i)} = \frac{P_{\mathrm{T}}^{(i)} V_{\mathrm{T}}\, M}{R_{\mathrm{u}} T_{\mathrm{H_2O}}} \ \text{and} \ \ m_{\mathrm{I}}^{(i)} = \frac{P_{\mathrm{I}}^{(i)} V_{\mathrm{I}}\, M}{R_{\mathrm{u}} T_{\mathrm{H_2O}}} \ . \tag{7}$$

The third mass component $m_{\mathrm{DE}}^{(i)}$ is the mass of gas that accumulates in the inventory volume during the "dead-end" time Δt_{DE} [4] when both *PVTt* diverter valves are closed at the start of a collection (< 100 ms). It also is calculated from values available from the *PVTt* system:

$$m_{\mathrm{DE}}^{(i)} = \dot{m}_{PVTt} \Delta t_{\mathrm{DE}} \ . \tag{8}$$

We used the approach described above to calculate temperature errors $T_{\mathrm{err}} = T - T_{\mathrm{H_2O}}$ for experiments conducted in the 34 L system at flows of 100 sccm, 500 sccm, 1 slm, 5 slm, and 10 slm and the results are plotted in Figure 6. Note that we intentionally applied the RoR method to larger flows than normal to better examine the temperature errors. Also, data were recorded at 10 Hz during these experiments (instead of the normally used 0.1 Hz) for better time resolution at the higher flows. For the 5 slm and 10 slm flows, critical flow venturis (CFVs) were used as the

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

meter under test. For flows of 5 slm or less, Molbloc* laminar flow meters were used as the meter under test. In general, meters under test are sensitive to pressure and to fluctuations of the flow. Therefore, RoR measurements are conducted with a back pressure regulator installed between the laminar flow meter and the collection tank. The back-pressure regulator maintains a steady pressure in the test section while the tank pressure increases during a fill. When the back-pressure regulator is installed (for flows ≤ 1 slm), the maximum collection pressure is 50 kPa rather than 100 kPa.

In Figure 6, there are two sets of data for 10 slm, one using a critical flow venturi (CFV) with throat diameter of 0.39 mm and an inlet pressure of 770 kPa and another using a 0.65 mm CFV and 295 kPa. There is no significant difference between the T_{err} traces for the two different CFVs.

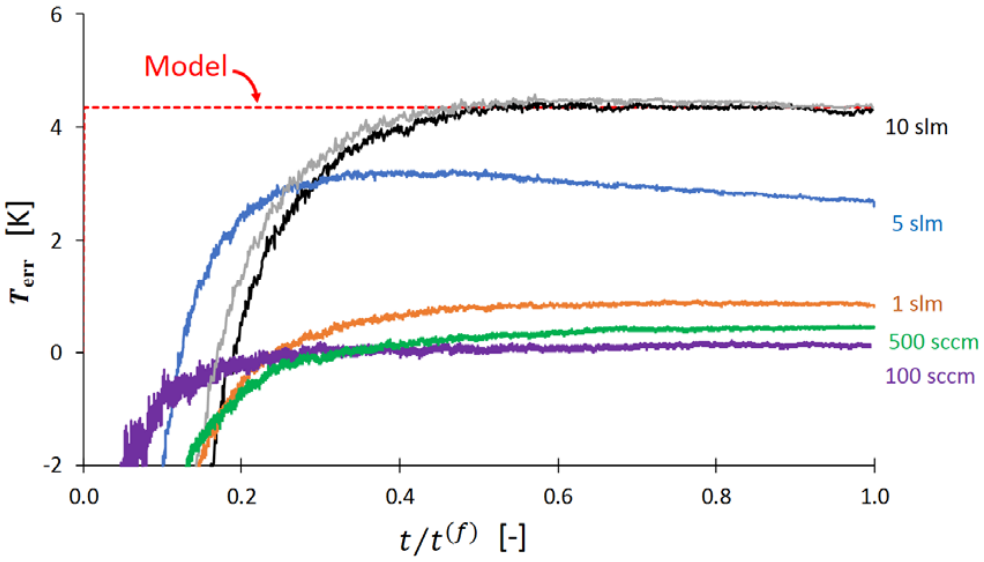


Figure 6. Temperature errors in the 34 L system during a RoR measurement for flows from 100 sccm to 10 slm. The data for 10 slm were taken with two critical flow venturis, one with a throat diameter of 0.39 mm and the other with a diameter of 0.65 mm.

Our lumped parameter thermodynamic model applies conservation of mass and energy to the filling tank [see Section 9 and reference 6]. It predicts that the temperature error is:

$$T_{\mathrm{err}} = T - T_{\mathrm{H_2O}} = T_{\mathrm{H_2O}} \left[\frac{(T_{\mathrm{in}}/T_{\mathrm{H_2O}})\gamma - 1}{1 + \Gamma} \right] \left[1 - \left(\frac{\dot{m} t}{m^{(i)}} + 1 \right)^{-(1+\Gamma)} \right] , \qquad (9)$$

where T is the gas temperature in the tank, T_{in} is the temperature of the gas entering the tank from the meter under test, γ is the ratio of the specific heats at constant pressure and constant

* Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

volume ($= c_P/c_V$), $\dot{m}$ is the mass flow, $m^{(i)}$ is the initial mass of gas in the tank before filling begins, and t is the time since the filling began. The dimensionless quantity Γ is defined by:

$$\Gamma \frac{(T-T_{\mathrm{H_2O}})}{T} = \left[\frac{hA}{c_V \dot{m}}\right]\left[\frac{(T-T_{\mathrm{H_2O}})}{T}\right], \tag{10}$$

where A is the surface area of the collection tank and h is the convective heat transfer coefficient between the gas and the tank walls. Equation 10 is the ratio of {heat transfer from the gas to its surroundings} to {the energy input to the gas by flow work}. The lumped-parameter analysis (Section 9) assumes that the gas is well stirred and that the tank wall temperature matches the water bath temperature, so that the heat transfer from the gas to the tank walls is the limiting heat transfer process.

Both the thermodynamic model and the pressure-equilibration data show that the temperature difference between the gas and the water bath reaches a steady state value during later portions of the fill and T_{err} is larger for larger flows (due to increased flow work). Unfortunately, values for the heat transfer coefficient are not readily available, but the model predictions for T_{err} using a value of h that forces agreement with the 10 slm steady state T_{err} are plotted in Figure 6. The model predicts that the temperature difference rapidly reaches a steady state value (< 0.2 s), but the experimental values in Figure 6 (based on pressure measurements made with a Heise HPO* sensor) show time constants of up to 8 min to reach steady state T_{err} values. We note that the lumped model does not account for the time required for convection to develop in the collection tank nor the transient response of the pressure sensor.

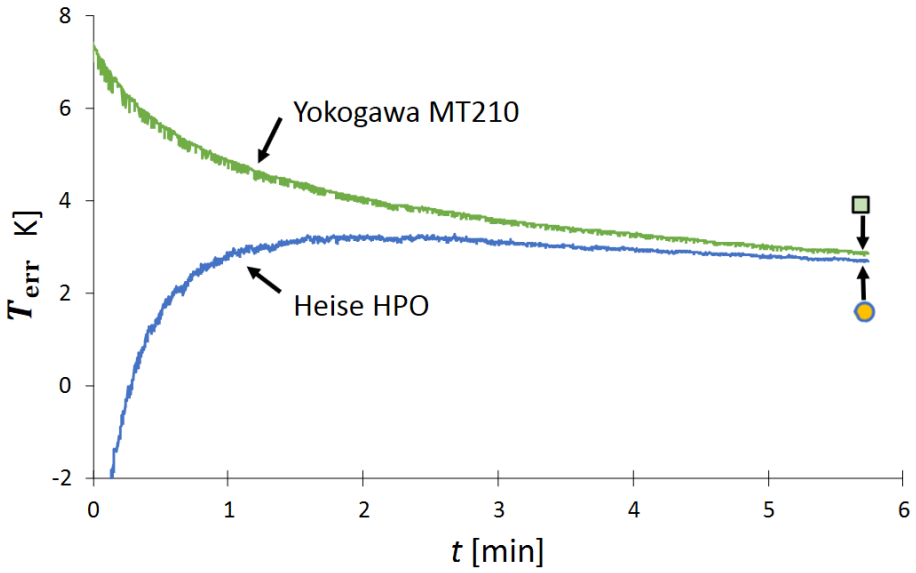


Figure 7. Temperature errors for 5 slm inferred from two sensors used to measure the tank pressure. The endpoints (indicated by symbols) are plotted for this and other flows in Figure 8.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Figure 7 implies that the transient response of the pressure sensor is important. The T_{err} traces based on two pressure sensors, the Heise HPO sensor used in Figure 6 and a Yokogawa MT210 pressure sensor, are quite different during earlier portions of the filling process. There are several phenomena that directly affect the time-dependence of the pressure sensor, such as 1) its mechanical and electrical response times, 2) the conductance of the tube connecting the sensor to the collection tank, and 3) flow-work-induced changes in the sensor's temperature. We hypothesize that the sensors in Figure 7 responded differently to these phenomena due to their different insulation or temperature control systems.

While Figure 7 shows that the pressure sensor's transient response can cause an error early in the fill process, it and analogous plots for other flows show that the T_{err} traces converge with the passage of time. Therefore RoR data from late in the fill process are less prone to errors caused by the response of pressure sensors to transients. Figure 8 plots T_{err} measured at the end of the collection *versus* flow: the plotted values correspond to the T_{err} values in Figure 6 where $t/t^{(f)} = 1$. Steady state values of T_{err} calculated from the Yokogawa MT210 pressure sensor agree well with those calculated from a Heise HPO pressure sensor.

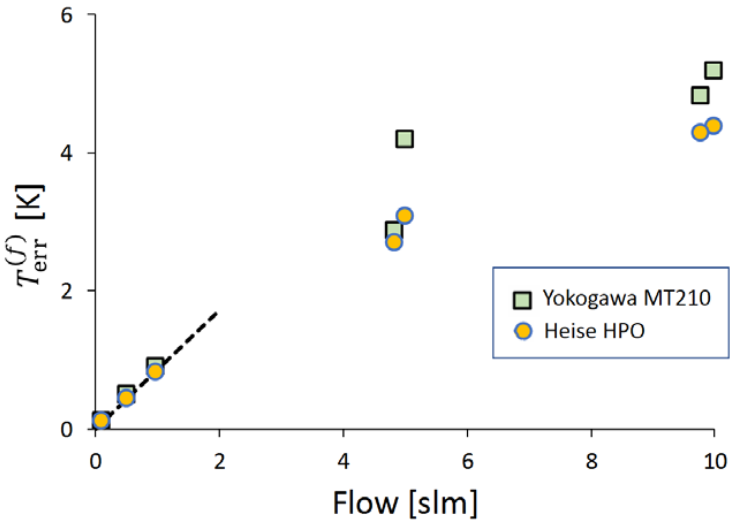


Figure 8. Difference between the gas temperature in the 34 L collection tank and the water bath temperature at the end of a RoR collection. The dashed line was used to produce the predicted RoR versus *PVTt* difference in Figure 3.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

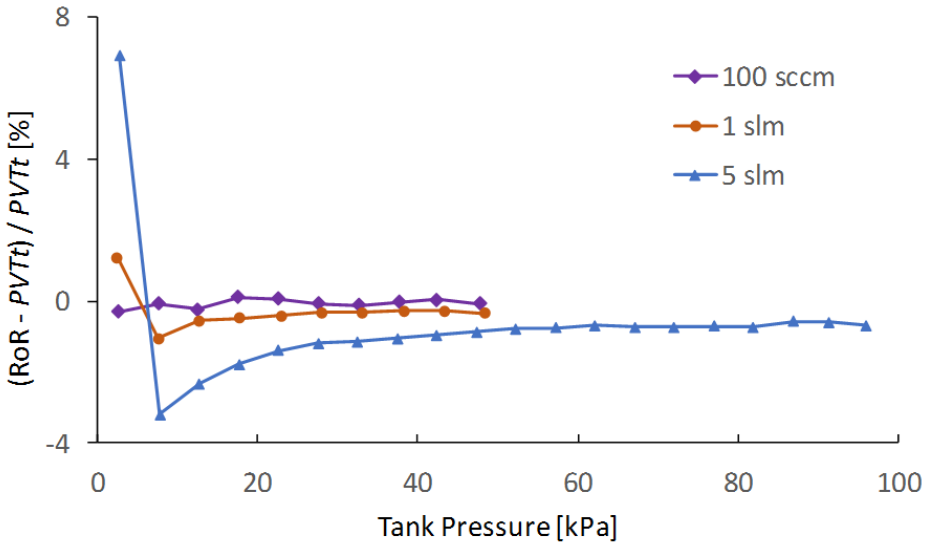


Figure 9. Difference between RoR and *PVTt* flow measurements using different portions of the data record, *i.e*. empty to 5 kPa, 5 kPa to 10 kPa, etc. The difference at high flows and low pressures is caused by pressure sensor response to transient conditions and pressure drop P_{err}.

Figure 9 compares RoR flow measurements for different portions of the filling process with the *PVTt* flow measurements. At lower tank pressures and larger flows, the errors in the RoR results are more than 5 %. As one would expect from Figure 6, as time passes and the tank reaches higher pressures, the RoR flow errors reach a steady value. Larger T_{err} and P_{err} values cause larger RoR flow uncertainties at larger flows (if uncorrected).

6. Uncertainty Analysis

The uncertainty of RoR gas flow measurements made using the NIST 34 L collection tank is plotted in Figure 10. The temperature and pressure errors T_{err} and P_{err} are treated as uncertainties (not corrections) in this analysis. For flows between 0.001 slm and 0.2 slm, the 34 L RoR flow standard has uncertainty between 0.05 % and 0.12 %.* For this range of flows, the dominant uncertainty varies between leaks, repeatability of the best existing device (BED), the collection volume, and density. At flows < 0.001 slm, the uncertainty is primarily due to the leak corrections and at flows > 0.2 slm, the RoR flow uncertainty is dominated by the uncorrected temperature errors T_{err}.

* All uncertainties herein are *k* = 2, approximately 95 % confidence level values.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

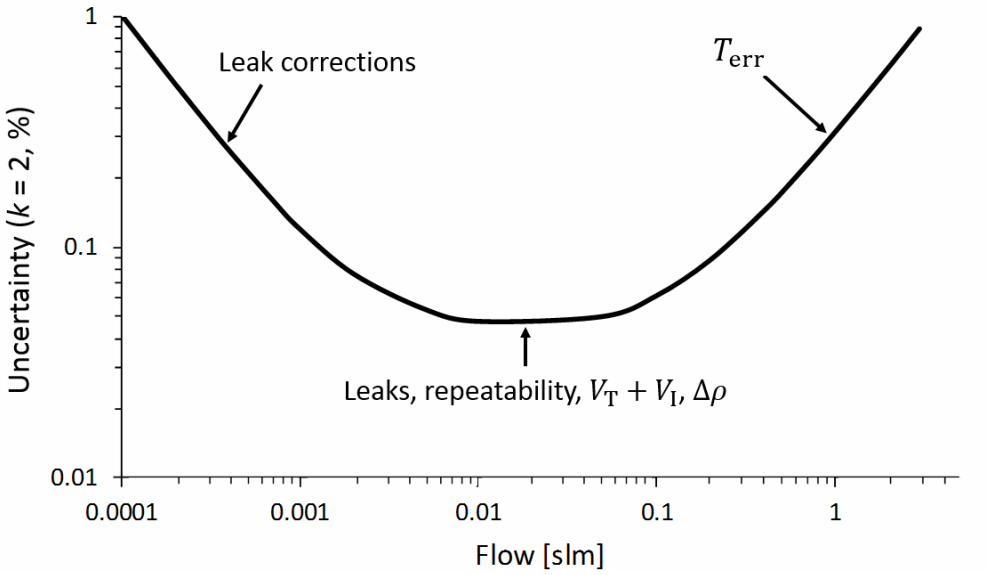


Figure 10. Uncertainty (*k* = 2, approximately 95 % confidence level) versus flow for the NIST 34 L Rate-of-rise flow standard.

The uncertainty components included in the analysis are:

1. Volume of the collection tank: $U(V_{\mathrm{T}} + V_{\mathrm{I}})$ is 9.7 cm^3 or 0.028 % [5].
2. Calibration, drift, and resolution of pressure, temperature, and time measurements: The pressure in the collection tank is measured with a pair of sensors that have an uncertainty of 0.006 kPa for pressures between 20 kPa and 130 kPa, are linear within 0.001 kPa over the same range, and have a resolution of 0.001 kPa. The RoR system is used at pressures greater than 20 kPa where the pressure sensors are known to be linear and transients in their readings due to the in-rush of gas (seen in Figures 6 and 7) have diminished to negligible levels, i.e. the time response of the pressure, temperature, and meter under test are not a concern due to proscribed operating conditions. The uncertainty of the water bath temperature is 0.012 K and the time uncertainty is 0.001 s.
3. Density calculations: The uncertainty of the compressibility factor (20 parts in 10^6) and molar mass (40 parts in 10^6) are included along with pressure and temperature uncertainties. The uncertainty of the universal gas constant is negligible (< 1 part in 10^6).
4. Leaks: The leak for the 34 L system has been measured for a range of tank pressures. The RoR flow measurements are corrected using a fit to the leak versus pressure data. The uncertainty of the leak correction is 5×10^{-4} sccm.
5. Repeatability of the BED is based on the standard deviation of the mean for calibration results from laminar flow meters.
6. Slope calculation: The uncertainty in calculation of slope is insignificant if data are collected for a sufficiently long interval. The uncertainty in the mass values m_j in Equation 1 is at

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

least 2.5 times larger than the uncertainty in the time values t_j, allowing us to apply the simplest expression for the expanded uncertainty of the slope calculation [2]:

$$U(a_1) = 2\left[\frac{\frac{\sum_{j=1}^{N}\left(m_j - a_1 t_j - a_0\right)^2}{N-2}}{\sum_{j=1}^{N} t_j^2 - \frac{\left(\sum_{j=1}^{N} t_j\right)^2}{N}}\right]^{\frac{1}{2}} = \frac{4\sqrt{3}\, s(m)}{\sqrt{N^3 - N}\ \Delta t}\ , \tag{11}$$

where a_0 and a_1 are the zeroeth and first order coefficients of the fit to the mass versus time data. (Here we have used a_1 instead of $\dot{m}_{\mathrm{RoR}}$ to avoid confusion: Equation 11 gives the uncertainty related to the fitting process, not the total uncertainty of the RoR mass flow.) The quantity $s(m)$ is the sample standard deviation of the mass fit residuals and Δt is the time interval between successive mass measurements. We used example data sets to quantify $s(m)$ for a range of flows and it led to < 0.01 % uncertainty contribution to the RoR mass flow.

For the NIST 34 L tank, a minimum of 1 h of 0.1 Hz data ($N = 360$ points) are processed to produce one RoR flow measurement. At flows below 1 sccm, data are collected for 18 h or more ($N \geq 6480$) so that the pressure change is ≥ 1 kPa in order to avoid problems due to the uncorrelated uncertainty of the pressure and temperature sensors. This is necessary because the normalized sensitivity coefficients $(x_i/\dot{m})(\partial\dot{m}/\partial x_i)$ for the pressures and temperatures used in the RoR calculation are inversely proportional to $P_N - P_1$. To illustrate the issue, consider the simplified case of a RoR measurement based on only two points, $m_1 = (V_{\mathrm{T}} + V_{\mathrm{I}})\,\rho(P_1, T)$ and $m_2 = (V_{\mathrm{T}} + V_{\mathrm{I}})\,\rho(P_2, T)$ separated by $\Delta t = t_2 - t_1$ so that $\dot{m}_{\mathrm{RoR}} = (m_2 - m_1)/\Delta t$. If the 95 % confidence level uncertainty due to non-linearity of the pressure sensor is 0.001 kPa, $P_1 = 20$ kPa, and $P_2 = 20.1$ kPa, this uncorrelated pressure uncertainty would lead to a mass flow uncertainty of 0.001 kPa / (20.1 kPa – 20 kPa) = 1 %. By filling the tank longer to increase the pressure change (and increase N), this uncertainty is reduced.

7. Uncertainty caused by flow instability is negligible. At NIST, we use a mass flow or pressure controller to ensure that the flow from the meter under test is stable within 0.2 % or better. Any data showing transients due to out-of-control environmental temperatures are removed prior to processing. Uncertainty due to flow instability is negligible because the meter under test has a short time response compared to the averaging interval (≥ 1 h) and because the RoR and meter under test data are averaged over the same interval. To visually qualify data that will be used to calculate the mass flow by Equation 1, we first plot the RoR flow calculated by:

$$\dot{m}_{\mathrm{RoR},j} \cong (V_{\mathrm{T}} + V_{\mathrm{I}})\left[\frac{\rho(P_{j+1}, T_{j+1}) - \rho(P_j, T_j)}{t_{j+1} - t_j}\right]\ , \tag{12}$$

The results from Equation 12 must be filtered (*e.g.* by a moving average of 10 or 100 points) to reduce noise from pressure resolution (Figure 11).

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

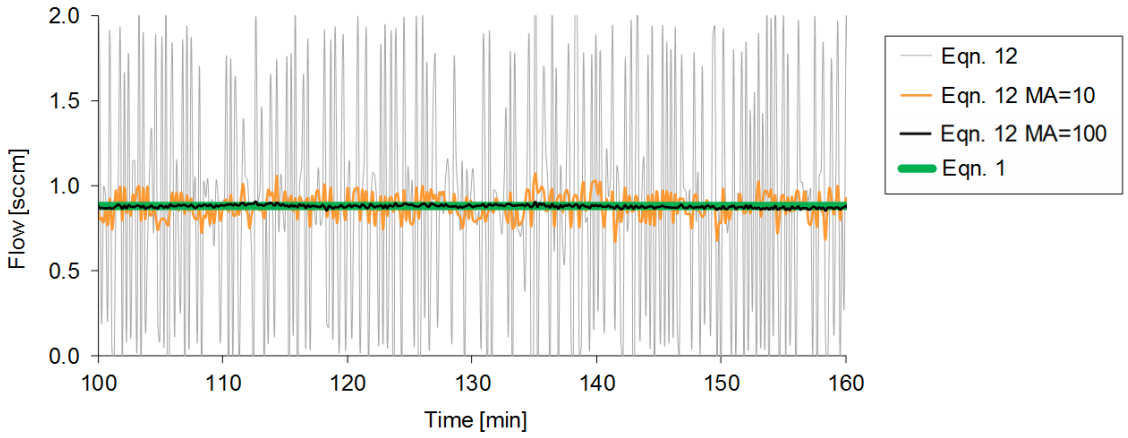


Figure 11. Example RoR flow from the best fit (Equation 1), and from Equation 12 with different moving average (MA) windows applied to check data for flow stability.

8. T_{err}, P_{err}: No corrections for the pressure drop between the pressure sensor and the collection tank P_{err} or for the temperature difference between the water bath and the gas T_{err} are made in the NIST RoR standard. At flows above 0.2 slm, the uncertainty increases because of P_{err}, T_{err}, and the decreasing number of data points N. The uncertainty at flows > 0.2 slm is predominantly due to T_{err}, not P_{err}, but the errors introduced by pressure drop can be significant for poorly selected operating conditions, i.e., if one filled the tank from empty to only 5 kPa at flows > 1 slm where the P_{err} and $\mathrm{d}P_{\mathrm{err}}/\mathrm{d}P$ are large (Figure 4).

Calculating the uncertainty due to the errors in temperature and pressure is complicated by the fact that T_{err} and P_{err} have asymmetric probability distributions and are correlated (both are tank pressure and flow dependent). They are asymmetric because we know that they always have positive values. In different circumstances, one might choose to treat T_{err} and P_{err} as corrections rather than errors or uncertainties. We have chosen not to use them as corrections in this case* because we do not intend to use RoR measurements for flows above 0.2 slm. (We will use *PVTt* measurements instead.) We emphasize the importance of quantifying these error sources. Methods for handling asymmetric uncertainty distributions are available [2], but for simplicity, we calculated the error in density change that would result from T_{err} and P_{err} for the particular flow and tank pressure conditions of interest and added that error to the root-sum-of-squares of all other components.

7. Advice for Designers and Operators of RoR Standards

A well designed RoR system will minimize the temperature rise due to flow work. This can be done by using a tank that has the characteristics of a heat exchanger: made of a thermally conductive material with large surface area. A water bath surrounding the collection tank is not strictly necessary; Nakao used a collection tank machined in a large piece of thermally conductive metal and measured the temperature of the metal instead of a water bath temperature [7]. The

* We use P_{meas} and $T_{\mathrm{H_2O}}$ to calculate density in Equation 1. In other circumstances, one would choose to use corrected pressure and temperature values ($P_{\mathrm{meas}} - P_{\mathrm{err}}$ and $T_{\mathrm{H_2O}} - T_{\mathrm{err}}$) instead.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

pressure sensors should be connected directly to the tank, not to the piping carrying the flow to the tank or any place where they are subject to flow (and pressure) dependent pressure drops.

Correlated pressure uncertainties (*e.g.*, a zero offset in calibration) do not contribute to RoR flow uncertainty, but gain errors or non-linearity of pressure sensors are a concern. Using the RoR system over a wider range of pressures reduces the uncertainty related to both pressure and temperature measurement. This can be done by increasing the fill time or reducing the collection volume. Therefore, a well-designed RoR standard has a collection tank that is small enough to produce large pressure changes in a short collection time, but has large surface area to improve heat transfer. NIST intends to build a new RoR standard to measure flows from 0.01 sccm to 1 sccm using a 0.3 L collection tank and fittings with low leak rates.

In this work, we benefited from having a validated *PVTt* standard to use as a reference to assess RoR flow and temperature measurements. But one can assess temperature corrections for flow work to a RoR standard without implementing *PVTt* measurements. The temperature of the gas at the end of a fill can be calculated using data like those in Figure 2 ($T = T_{\mathrm{H_2O}}\, P^{(f)}/P_{eq}^{(f)}$). Temperature errors can be tabulated as a function of flow and an empirical correction applied to $T_{\mathrm{H_2O}}$ to improve flow measurements at high flows where temperature errors are a large uncertainty source. The uncertainty analysis for this approach would need to include the uncertainty of the corrections. In lieu of making corrections, one can use a collection tank with greater heat transfer to the surroundings (>hA) to reduce T_{err} to small values.

The comparison of RoR and *PVTt* data revealed that there are transient effects, apparently related to the in-rush of gas at the beginning of a collection, that last much longer than the expected response time of the sensors to pressure changes. We speculate that the long transients are related to the internal temperature of the pressure transducers. Sensors don't necessarily behave with a first order time response and a single time constant. A pressure sensor exposed to a step change in pressure may have a short first-order time response of large magnitude for electrical or mechanical reasons, but also have a longer second-order response of small magnitude for thermal reasons. The transient response had opposite signs for the two types of pressure sensors we used, but both types performed well in later portions of a fill (P > 20 kPa). These problems can be avoided by making plots like Figure 9 (using the RoR data over the full available pressure range instead of *PVTt* data) and only using results that indicate that steady state has been attained.

8. Acknowledgements

The authors are grateful to Iosif Shinder and Robert Berg for their advice on slope calculation uncertainty and many other topics.

9. Appendix: Derivation of the Temperature Error Due to Flow Work (Equation 9)

Assuming 1) the thermodynamic properties of the gas are uniform in the control volume defined by the tank internal walls (a "lumped parameter analysis"), 2) there is no work other that flow work applied, 3) energy is conserved, and 4) the gas is ideal and has constant specific heat, Equation 11 of reference [6] with heat transfer gives:

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$c_V m \frac{\mathrm{d}T}{\mathrm{d}t} + c_V T \frac{\mathrm{d}m}{\mathrm{d}t} = \dot{m} c_P T_{\mathrm{in}} - hA(T - T_{\mathrm{s}}) \ , \tag{A1}$$

where T is the temperature of the gas in the collection tank, T_{in} is the temperature of the gas entering the tank, T_{s} is the temperature of the surface of the tank, m is the mass of gas in the tank at time t, $\dot{m} = \mathrm{d}m/\mathrm{d}t$ is the mass flow of gas entering the tank, c_V and c_P are the gas constant volume and constant pressure specific heat capacities respectively, h is the convective heat transfer coefficient between the gas and the tank walls, and A is the surface area of the tank walls. Equation A1 can be rewritten as:

$$c_V m \frac{\mathrm{d}(T-T_{\mathrm{s}})}{\mathrm{d}t} + c_V (T - T_{\mathrm{s}}) \frac{\mathrm{d}m}{\mathrm{d}t} + hA(T - T_{\mathrm{s}}) = \dot{m} c_P T_{\mathrm{in}} - c_V \frac{\mathrm{d}m}{\mathrm{d}t} T_{\mathrm{s}} \ . \tag{A2}$$

If mass flow $\dot{m}$ is diverted into the tank at $t = 0$,

$$m(t) \ = m^{(i)} + \dot{m} t \ , \tag{A3}$$

and this expression can be substituted into Equation A2 to give:

$$c_V \left[\dot{m} t + m^{(i)}\right] \frac{\mathrm{d}(T-T_{\mathrm{s}})}{\mathrm{d}t} + [c_V \dot{m} + hA](T - T_{\mathrm{s}}) = \dot{m}[c_P T_{\mathrm{in}} - c_V T_{\mathrm{s}}] \ . \tag{A4}$$

Dividing by c_V results in:

$$\left[\dot{m} t + m^{(i)}\right] \frac{\mathrm{d}(T-T_{\mathrm{s}})}{\mathrm{d}t} + \left[\dot{m} + \frac{hA}{c_V}\right](T - T_{\mathrm{s}}) = \dot{m}[T_{\mathrm{in}}\gamma - T_{\mathrm{s}}] \ \ , \tag{A5}$$

where $\gamma = c_P/c_V$. Introducing the dimensionless variables $\hat{t} = t/t^{(f)}$, $\hat{T}_{\mathrm{in}} = T_{\mathrm{in}}/T_{\mathrm{s}}$, $\hat{\theta} = (T - T_{\mathrm{s}})/T_{\mathrm{s}}$, and $\Gamma = \frac{hA}{c_V \dot{m}}$ transforms Equation A5 to:

$$\left[\hat{t} + \frac{m^{(i)}}{\dot{m}\, t^{(f)}}\right] \frac{\mathrm{d}\hat{\theta}}{\mathrm{d}\hat{t}} + [1 + \Gamma]\hat{\theta} = \hat{T}_{\mathrm{in}}\gamma - 1 \ \ , \tag{A6}$$

which for the boundary condition $T = T_{\mathrm{s}}$ at $t = 0$ (*i.e.* $\hat{\theta}(0) = 0$) has the solution:

$$\hat{\theta} = \left(\frac{\hat{T}_{\mathrm{in}}\gamma - 1}{1 + \Gamma}\right)\left[1 - \left(\frac{\dot{m}\, t^{(f)}}{m^{(i)}}\hat{t} + 1\right)^{-(1+\Gamma)}\right], \tag{A7}$$

for $\hat{t} < 1$. Converting back to the dimensional variables and assuming that the surface temperature of the tank equals the water bath temperature $T_{\mathrm{s}} = T_{\mathrm{H_2O}}$ gives:

$$T_{\mathrm{err}} \ = \ T - T_{\mathrm{H_2O}} = T_{\mathrm{H_2O}}\left[\frac{(T_{\mathrm{in}}/T_{\mathrm{H_2O}})\gamma - 1}{1 + \Gamma}\right]\left[1 - \left(\frac{\dot{m} t}{m^{(i)}} + 1\right)^{-(1+\Gamma)}\right] \ . \tag{A8}$$

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

[1] Lemmon, E. W., Huber, M. L., and McLinden, M. O., *Refprop 23: Reference Fluid Thermodynamic and Transport Properties*, NIST Standard Reference Database 23, Version 9, National Institute of Standards and Technology, Boulder, CO, November, 2010.

[2] Coleman, H. W., and Steele, W. G., *Experimentation, Validation, and Uncertainty Analysis for Engineers*, 3rd edition, A. John Wiley and Sons, 2009, Appendix E.

[3] *Method for Vacuum Leak Calibration*, American Vacuum Society, AVS Standard 2.2-1968, Vac. Sci. Tech., Vol. 5, pp. 219-222, 1968.

[4] Wright, J. D., Johnson, A. N., Moldover, M. R., and Kline, G. M., *Gas Flowmeter Calibrations with the 34 L and 677 L PVTt Standards*, NIST Special Publication 250-63, National Institute of Standards and Technology, Gaithersburg, Maryland, November 18, 2010, http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication250-63.pdf.

[5] Wright, J. D. and Johnson, A. N., *NIST Lowers Gas Flow Uncertainties to 0.025 % or Less*, NCSL International Measure, 5, pp. 30 to 39, March, 2010, http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=901413.

[6] Wright, J. D. and Johnson, A. N., *Uncertainty in Primary Gas Flow Standards Due to Flow Work Phenomena*, FLOMEKO, Salvador, Brazil 2000, https://www.nist.gov/sites/default/files/documents/calibrations/flomeko_2000.pdf.

[7] Nakao, S.-I., *Development of the PVTt System for Very Low Gas Flow Rates*, Flow Measurement and Instrumentation, 17, pp. 193 to 200, 2006.

Johnson, Aaron; Kline, Gina; Moldover, Michael; Wright, John.
"Errors in Rate-of-Rise Gas Flow Measurements from Flow Work."
Paper presented at International Symposium for Fluid Flow Measurement, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Proceedings of the ASME 2018 International Manufacturing Science and Engineering Conference
MSEC2018
June 18 - 22, 2018, College Station, Texas, USA

MSEC2018-6612

DESIGN OF AN OCTO-STRAIN SPECIMEN FOR BIAXIAL TENSION TESTING

Justin L. Milner and Thomas Gnäupel-Herold
National Institute of Standards and Technology,
NIST Center for Neutron Research
Gaithersburg, MD, USA

KEYWORDS
Sheet metal, Biaxial testing, Cruciform specimen, Octo-Strain specimen, Digital image correlation

ABSTRACT
A custom biaxial testing fixture was designed to evaluate a new specimen geometry for complex loading paths. Biaxial testing is commonly used to evaluate work-hardening behavior of sheet metal in biaxial tension to study the accumulation of plastic strains to determine the anisotropic yield loci. The current state-of-the-art specimen geometry that is used for biaxial testing is the cruciform specimen. Cruciform specimens are machined into a geometry that resembles a cross with four arms arranged at 90 degrees. However, this geometry is prone to premature failure and non-homogenous strain distribution within the gauge region. These problems persist even with the addition of complex features (*e.g.,* slits and multi-step pockets). Therefore, the primary goal of the new specimen geometry is to achieve a large and uniform strain field within the gauge region. One of the main problems of the cruciform specimen is the formation of stress concentration within the gauge region. Therefore, the proposed specimen geometry is comprised of four additional arms between the existing cruciform arms. This geometry is termed an 'Octo-Strain' specimen after the eight arms that are arranged in a 45-degree planar pattern. It is hypothesized that the additional arms will stabilize the stress concentrations and, thus, achieve increased failure strain and uniformity as compared to the cruciform geometry. This work focuses on the comparison of the cruciform specimen to the Octo-Strain specimen during balanced biaxial deformation of mild steel. It is found that the Octo-Strain specimen achieved twice the failure strain and increased strain uniformity within the gauge region as compared to the cruciform specimen.

INTRODUCTION
Biaxial testing is a method to determine material data that can be used for finite element modeling by deforming sheet metal material under complex strain paths. These methods include punch tests (*e.g.,* Marciniak and Nakazima) [1-3], bulge pressure tests [4-6], biaxial compression tests [7], and cruciform tests [8-10] to name a few. Each of these techniques has their advantages and disadvantages; such cons include the introduction of bending stresses and friction during the test, sample geometry issues, and measuring the stress-strain relationship. Cruciform testing eliminates the issues associated with the introduction of bending stresses and friction as it is an in-plane testing technique. The cruciform specimen geometry is derived from the standard uniaxial tensile test specimen, with the addition of a second loading direction resulting in a geometry that resembles two uniaxial specimens arranged perpendicular to each other (resembling a cross). The arms of the cruciform specimen are given a displacement that induces a biaxial force over the planar gauge area (at the intersection of the arms or center of the specimen). However, the cruciform specimen geometry tends to lead to premature failure and, thus, the failure strain limit of the material is not reached. This is primarily due to high-stress localization that occurs within the fillet area of the specimen near the radius (connecting the arms). Therefore, a radically new specimen geometry is considered to replace the cruciform specimen.

This new geometry is an eight-arm specimen arranged in a 45-degree planar pattern or, in other words, two cruciform specimens with a 45-degree relative offset. It is termed the Octo-Strain specimen due to its eight arms. This geometry was chosen to stabilize the stress localization seen to occur with the cruciform specimen geometry. It is also derived from the punch test specimen geometries which have a 360-degree clamping

Gnaupel-Herold, Thomas; Milner, Justin L.
"Design of an Octo-Strain Specimen for Biaxial Tension Testing."
Paper presented at ASME 2018 International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

force around the sample. More arms will approximate 360° of clamping. However, space requirements are the limiting factor for the arrangement of 2^n arms, where n is greater than 3. Similar to the cruciform specimen, a geometry with uniform thickness and arm width is insufficient for plastic deformation in the gauge region. Therefore, a reduction in thickness (recess/pocket) is required to further localize plastic deformation within the gauge region. Figure 1 displays a schematic of the Octo-Strain specimen geometry with key definitions of dimensions. This geometry follows the same necessities as the cruciform specimen in that a central recess (gauge region) of reduced cross-section is needed to localize plastic deformation. The criteria that the Octo-Strain specimen should satisfy is (1) Uniform strain distribution within the gauge region. (2) Large strains are achievable in the gauge region. (3) Failure should occur within the gauge region. (4) Adequate gauge area for stress and strain measurement. (5) And the specimen design should be simple and economical to manufacture.

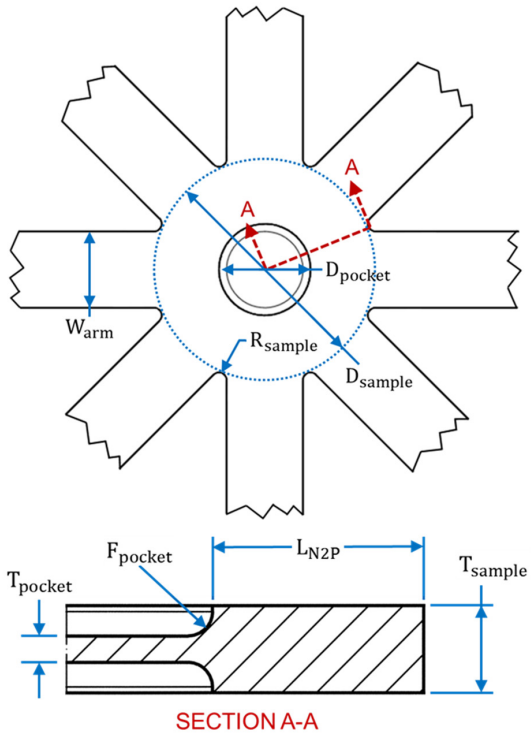


Figure 1: Schematic of an Octo-Strain specimen geometry and definitions of dimensions.

With a new Octo-Strain specimen geometry comes a new design of the biaxial testing system, the Octo-Strain electro-mechanical frame. Octo-Strain, shown in Figure 2, was custom designed to test the eight-arm samples under complex loading paths with in-situ stress and strain measurements via neutron diffraction and digital image correlation, respectively. This loading frame has eight independently controlled actuators with a capacity of ±15 kN and a stroke of ~30 mm. Each actuator can be controlled in displacement, load or strain control mode with the ability to switch between modes without unloading. Numerous strain paths can be achieved through varying the actuator velocities. The deformation of the samples is recorded using full-field digital image correlation (DIC), and this system can also be used for real-time strain control. Furthermore, in-situ neutron diffraction for stress measurement is intended during testing but is not used for the tests described in this paper. However, this capability leads to unique design criteria resulting in an "open" design to not obscure the neutron beam path and 360-degree rotation about the sample normal allowing for the determination of the planar stress tensor components.

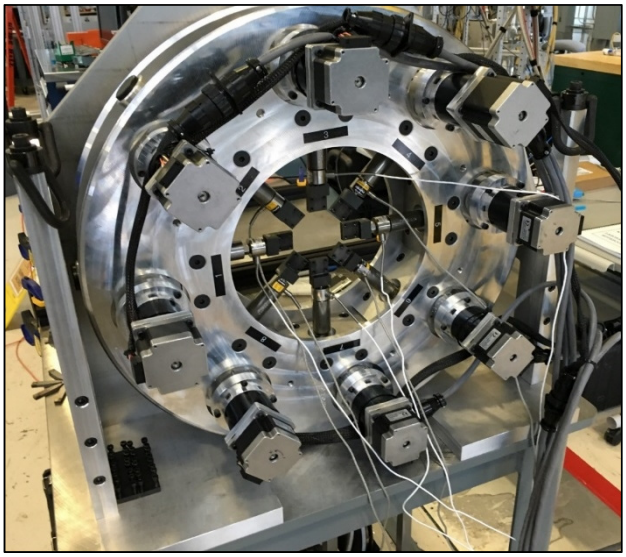


Figure 2: Photograph of the Octo-Strain biaxial loading device.

This work focuses on the comparison between the cruciform specimen geometry and the new Octo-Strain specimen geometry under balanced biaxial loading. Full-field strain mapping evaluates the performance of the specimen geometries tested via digital image correlation.

EXPERIMENTAL PROCEDURE

Experiments are conducted to evaluate the performance of the Octo-Strain specimen and compared to the cruciform specimen. The material used in this study is cold-rolled mild steel (AISI 1010) sheets received in a thickness of 3.31 ± 0.01 mm. All samples were waterjet cut along the outside profile, and the sample pocket was machined with a 1 mm corner radius end mill. A custom designed jig was created to position the samples, during the machining of the pocket, to ensure a centered pocket with respect to the sample. The experiments were performed with the Octo-Strain biaxial testing system (shown in Figure 2),

Gnaupel-Herold, Thomas; Milner, Justin L.
"Design of an Octo-Strain Specimen for Biaxial Tension Testing."
Paper presented at ASME 2018 International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

as described earlier. The specimens were tested under balanced biaxial deformation using load control to maintain a constant load ratio of one. The tests were conducted until the specimen fractured or when necking was observed within the arms. The strain evolution of the specimens was recorded with a full-field digital image correlation system.

To evaluate the samples equally, peak force was taken as a reference point for each test. The peak force was determined by taking the average of the load cell readings from the eight arms (four for the cruciform sample), then determining the maximum value. At peak force, a subset of the gauge region is analyzed to determine the average strain and strain homogeneity. The gauge region of the sample is defined as $D_{pocket} - 2 \times F_{pocket}$, which is the flat section of the sample pocket. The area used for determining strain and strain homogeneity it termed 'gauge area' and is defined as the area within a diameter equal to $0.3 \times D_{pocket}$. Furthermore, strain values along four line slices (two line slices for the cruciform geometry) connecting the sample's radii was extracted and averaged into a single line slice (see Figure 3), to analyze the strain profile. The gauge area and line slice is schematically shown in Figure 3.

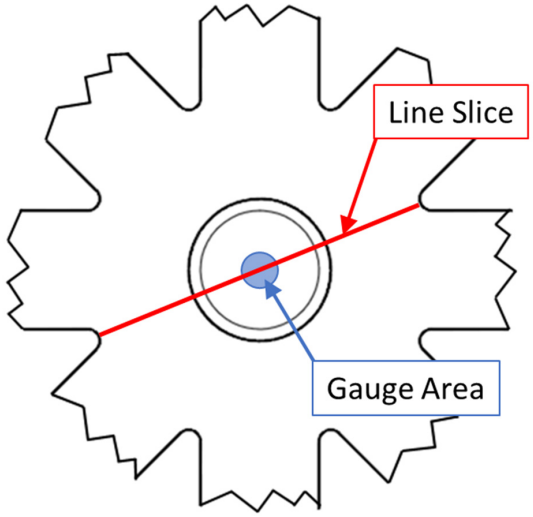


Figure 3: Schematic representation of the analyzed gauge region and line slice.

RESULTS AND DISCUSSION

To evaluate the performance of Octo-Strain specimen and compare it to the cruciform specimen without bias, geometries where chosen to compare the two best. Preliminary testing was conducted on the two specimen geometries, and it was determined that decreasing the thickness ratio, defined as the pocket thickness divided by the sample thickness ($t_{ratio} = t_{pocket}/t_{sample}$), increases the strain achieved in the gauge region regardless of other key parameters (*i.e.* pocket diameter). Therefore, the thickness ratio of 0.25 was chosen for testing both specimen geometries, this corresponds to a $t_{sample} = 3.2\ mm$ and $t_{pocket} = 0.8\ mm$. Multiple sample radii (R_{sample}) were tested for both the specimen geometries and it was determined that a R_{sample} = 12 mm (cruciform) and a R_{sample} = 1 mm (Octo-Strain) provided the best results. These specimen geometries discussed in detail in this work are not considered optimal, as it may require more complex geometries of the specimen and/or gauge region. Therefore, in this work, only the pocket diameter (D_{pocket}) was evaluated experimentally to determine optimal performance, then the optimal specimen is chosen for further examination and comparison.

For the cruciform specimen, the D_{pocket} was varied between 15 mm and 18 mm at 1 mm intervals. These results are portrayed in Figure 4, displaying resultant pocket diameter verses effective strain at peak force. Decreasing the pocket diameter increases the strain achieved within the gauge region, until a drastic drop off in the strain level is noticed at a diameter of 15 mm. This reduction in strain is due to the increased sample material around the gauge region causing stain localization to occur in the arms of the specimen. This occurs around a D_{pocket} of 15 mm for the cruciform specimen with a R_{sample} of 12 mm, thus a D_{pocket} of 16 mm is near optimal for this geometry and selected for further discussion.

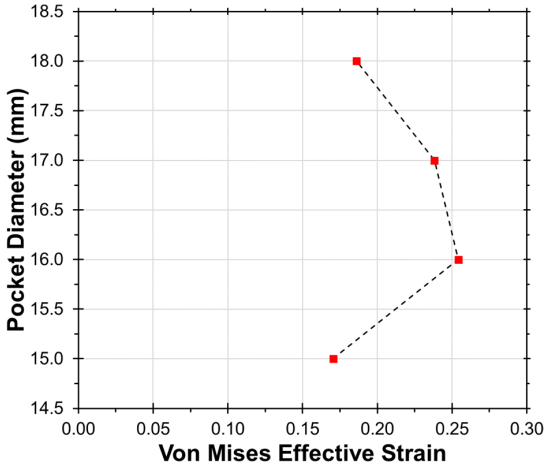


Figure 4: Pocket Diameter verse effective strain of cruciform sample geometry.

Figure 5 displays (a) average force per arm verse effective strain and (b) strain path of the cruciform specimen with a D_{pocket} of 16 mm. The horizontal and vertical dashed lines, within Figure 5, indicates peak force and corresponding effective strain. Peak force was recorded at 8.97±0.12 kN occurring at an effective strain of 0.247±0.007. The horizontal and vertical dashed lines in Figure 5(b) represents the strain achieved in the x-direction (ε_{xx}) and y-direction (ε_{yy}) at peak force. The strain path is nearly ideal balanced biaxial where ε_{xx}= 0.123±0.008 and

Gnaupel-Herold, Thomas; Milner, Justin L.
"Design of an Octo-Strain Specimen for Biaxial Tension Testing."
Paper presented at ASME 2018 International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

ε_{yy}= 0.123±0.006 at peak force. Furthermore, the strain path remains balanced biaxial just prior to failure. It should be noted that the acquisition rate was changed mid-test from 1 frame per 2 seconds to 1 frame per 4 seconds without stopping deformation. This is most notable in Figure 5(b), where a slight pause is seen around a balanced strain of 0.11 and the data rate appears to slow thereafter. This change in the acquisition rate does not affect the test results.

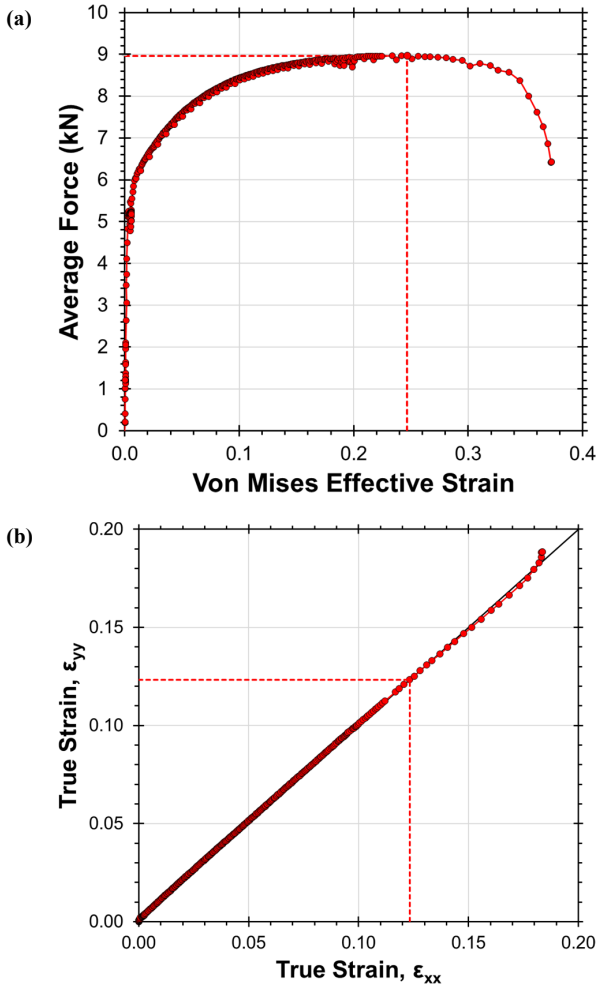


Figure 5: (a) average force verse effective strain and (b) strain path of the cruciform specimen with 16 mm pocket diameter

To further investigate the performance of the cruciform specimen geometry, the full-field strain data was examined for strain homogeneity and strain profile. Figure 6(a) displays the full-field strain map overlaid on the deformed sample at peak force. The strain is seen to localize near the pocket fillet at the location closest to the sample radius. This is more evident by the strain profile, shown in Figure 6(b), represented by the solid red line. This profile is derived from the line slice data and is plotted based on the undeformed sample. The break in the strain profile data is where DIC is unable to determine strains due to geometry effects caused by the pocket fillet. The cruciform specimen exhibits a maximum effective strain of 0.47 near the pocket fillet, as compared to the minimum of 0.24 at the center of the gauge region. This significant variation of strain within the gauge region results in a less homogenous strain distribution. Analyzing the cruciform specimen results in an effective strain of 0.247 with a 0.007 standard deviation. This is represented graphically in Figure 6(b) where the blue lines depict the gauge area. The solid line signifies the mean and the dashed lines indicate the upper and lower bounds. The width of the lines also implies the diameter of the circular gauge area.

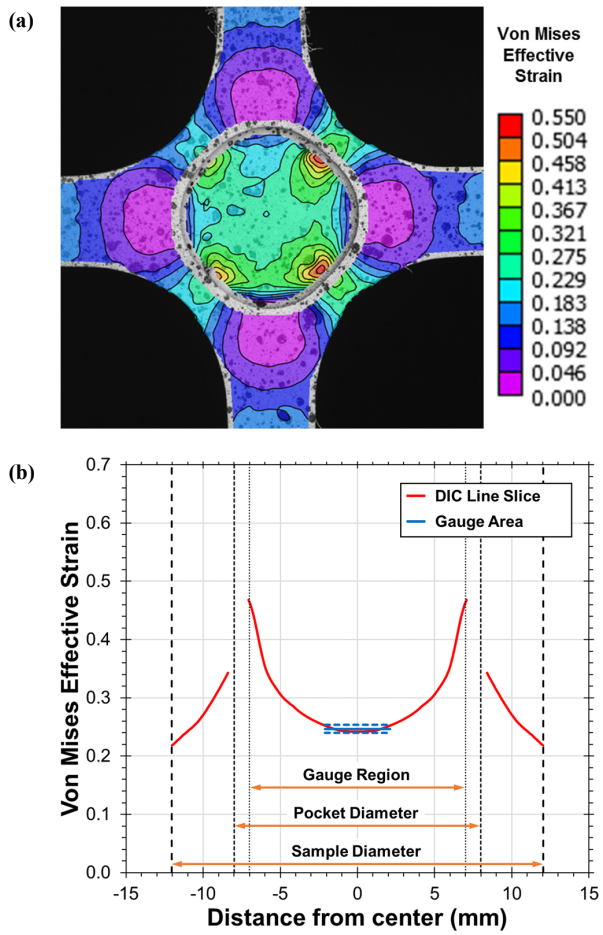


Figure 6: Cruciform sample showing (a) DIC measured von Mises effective strain (at peak force) overlaid on the image and (b) effective strain profile along line slice (red line) with average gauge area strain and variation (blue lines).

These results of the cruciform specimen compare favorably with the work of Creuziger et al. [11], where they determined an optimized cruciform geometry through FEA. Their optimized

Gnaupel-Herold, Thomas; Milner, Justin L.
"Design of an Octo-Strain Specimen for Biaxial Tension Testing."
Paper presented at ASME 2018 International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

specimen, of mild steel, resulted in an effective strain of 0.246±0.011 determined at 90% of the strain at peak force and within a 25% gauge area. Using the same definitions as described in [11], the cruciform specimen with a pocket diameter of 16 mm, analyzed here, results in an effective strain of 0.227±0.011. In comparison, the cruciform geometry in this work achieved 0.019 less strain as compared to [11], corresponding to an 8% difference. Again, the cruciform geometry tested in this work is not considered optimized.

For the Octo-Strain specimen, the results of varying the pocket diameter is portrayed in Figure 7, where the D_{pocket} was varied between 10 mm and 14 mm with an interval of 1 mm while skipping 13 mm. These results exhibited a similar trend as observed with the cruciform specimen geometry, where decreasing the pocket diameter increases the strain achieved within the gauge region until a drastic drop is noticed. This drop occurs when strain localizes within the arms of the Octo-Strain specimen, around a pocket diameter of 10 mm. The Octo-Strain specimen with a pocket diameter of 12 mm slightly outperformed the specimen with 11 mm pocket diameter. It is assumed that the optimal geometry of the Octo-Strain specimen is near the 11 mm pocket diameter. This is supported by FE modeling that will be discussed in detail in a later paper. This occurrence may be due to slight thickness variation within the gauge region. Therefore, the Octo-Strain specimen with a pocket diameter of 12 mm is selected for further examination and comparison.

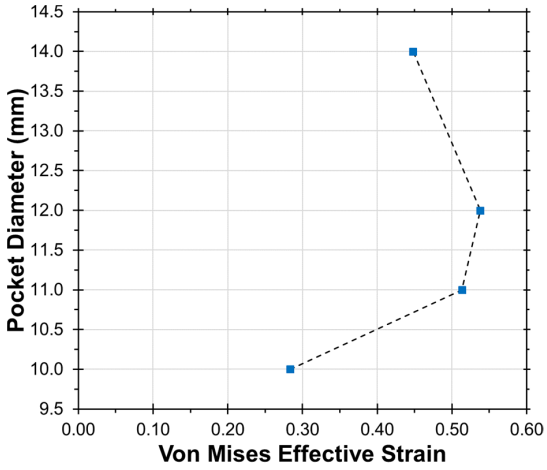


Figure 7: Pocket Diameter verse effective strain of Octo-Strain sample geometry.

The average force per arm verse effective strain and the resultant strain path for the Octo-Strain specimen with a pocket diameter of 12 mm are portrayed in Figure 8, along with the results of the cruciform specimen in gray for reference. The average peak force per arm is 9.07±0.12 kN, which is similar to the cruciform specimen. At peak force, the effective strain was 0.538±0.003, which is over twice the strain achieved by the cruciform specimen. This is further realized by the strain path where ε_{xx}= 0.262±0.002 and ε_{yy}= 0.275±0.002 at peak force. The strain path of the Octo-Strain specimen veers off balanced biaxial at higher strains. This is due to the load control regime that was chosen for these tests, and if strain control was employed this deviation would not arise. The load deviation throughout the test was 121 N, even up to failure.

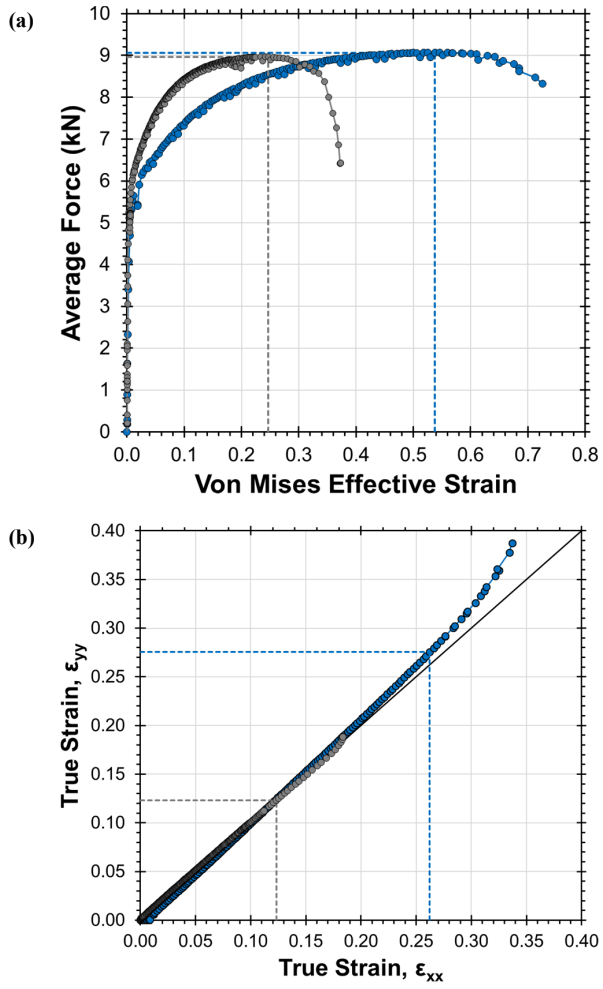


Figure 8: (a) average force verse effective strain and (b) strain path of the Octo-Strain specimen with 12 mm pocket diameter. The cruciform specimen with 16 mm pocket diameter is also shown in gray for reference.

To further compare the Octo-Strain specimen with the cruciform specimen, the full-field DIC data is examined and displayed in Figure 9. The maximum effective strain of 0.57 is seen to occur at the samples' radii, as observed in Figure 9(a). Moreover, the gauge region of the Octo-Strain specimen has a more homogeneous strain field. This is more evident when examining the strain profile, shown in Figure 9(b), represented

Gnaupel-Herold, Thomas; Milner, Justin L.
"Design of an Octo-Strain Specimen for Biaxial Tension Testing."
Paper presented at ASME 2018 International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

by the solid red line. A large strain gradient occurs from the sample radius to the pocket diameter by a factor of three. Nevertheless, the strain gradient within the gauge region is minor in comparison, with a rather uniform strain field in the center of the gauge region. This provides a sufficient area (approximately 6 mm in diameter) for measuring stress, via neutron diffraction, with minimal error due to strain uncertainty. Analyzing the Octo-Strain specimen results in a measured effective strain of 0.538±0.003, graphically portrayed in Figure 9. The gauge area is depicted by the blue lines, where the solid line indicates the average effective strain and the dashed lines represent the variation.

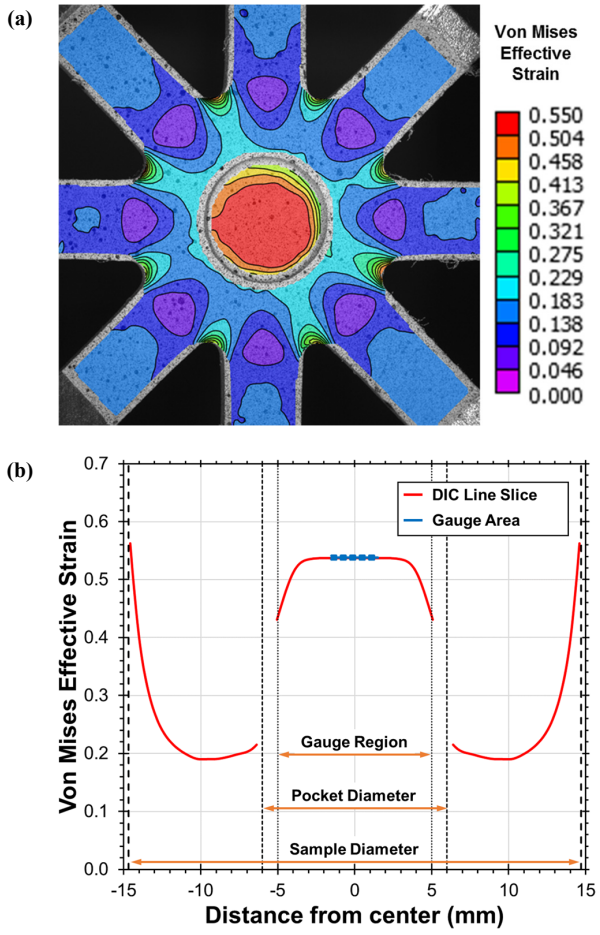


Figure 9: Octo-Strain sample showing (a) DIC measured von Mises effective strain (at peak force) overlaid on the image and (b) effective strain profile along line slice (red line) with average gauge area strain and variation (blue lines).

SUMMARY AND CONCLUSIONS

The findings presented here are the first glimpse into the design and performance of the Octo-Strain specimen, a novel geometry for biaxial testing. When comparing the two specimen geometries, the Octo-Strain geometry outperforms the cruciform. In terms of strain at peak force, the cruciform reached an effective strain of 0.247, as compared to 0.538 for the Octo-Strain specimen. This corresponds to a twofold increase in strain achieved for the Octo-Strain specimen. Furthermore, the strain field within the gauge region is more homogeneous for the Octo-Strain specimen, leading to more accurate measurements. This homogeneity increase is due to a smaller strain gradient within the gauge region as compared to the cruciform specimen. Future work will consist of FEA modeling to determine an optimized Octo-Strain geometry and testing of various materials.

REFERENCES

[1] Z. Marciniak and K. Kuczyński, "Limit Strains in the Processes of Stretch-Forming Sheet Metal," *International Journal of Mechanical Sciences,* vol. 9, pp. 609-620, 1967.

[2] K. Nakazima, T. Kikuma, and K. Hasuka, "Study on the Formability of Steel Sheets," *Yawata Technical Report,* vol. 264, pp. 8517-8530, 1968.

[3] S. P. Keeler and W. A. Backofen, "Plastic Instability and Fracture in Sheets Stretched Over Rigid Punches," *ASM Trans Q,* vol. 56, pp. 25-48, 1963.

[4] A. Ranta-Eskola, "Use of the Hydraulic Bulge Test in Biaxial Tensile Testing," *International Journal of Mechanical Sciences,* vol. 21, pp. 457-465, 1979.

[5] D. Rees, "Plastic Flow in the Elliptical Bulge Test," *International Journal of Mechanical Sciences,* vol. 37, pp. 373-389, 1995.

[6] J. Liu, M. Ahmetoglu, and T. Altan, "Evaluation of Sheet Metal Formability, Viscous Pressure Forming (VPF) Dome Test," *Journal of Materials Processing Technology,* vol. 98, pp. 1-6, 2000.

[7] Y. Tozawa and M. Nakamura, "New Method for Testing Sheet Under Biaxial Compression," *Journal of Japan Society for Technology of Plasticity,* vol. 13, pp. 538-541, 1972.

[8] E. Shiratori and K. Ikegami, "Experimental Study of the Subsequent Yield Surface by Using Cross-Shaped Specimens," *Journal of the Mechanics and Physics of Solids,* vol. 16, pp. 373-394, 1968.

[9] S. Demmerle and J. P. Boehler, "Optimal Design of Biaxial Tensile Cruciform Specimens," *Journal of the Mechanics and Physics of Solids,* vol. 41, pp. 143-181, 1993.

[10] W. Müller and K. Pöhlandt, "New Experiments for Determining Yield Loci of Sheet Metal," *Journal of Materials Processing Technology,* vol. 60, pp. 643-648, 1996.

[11] A. Creuziger, M. A. Iadicola, T. Foecke, E. Rust, and D. Banerjee, "Insights into Cruciform Sample Design," *JOM,* vol. 69, pp. 902-906, 2017.

Gnaupel-Herold, Thomas; Milner, Justin L.
"Design of an Octo-Strain Specimen for Biaxial Tension Testing."
Paper presented at ASME 2018 International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

Uncertainty of the Ohm Using Cryogenic and Non-Cryogenic Bridges

Alireza R. Panna[*], Marlin E. Kraft[*], Albert F. Rigosi[*], George R. Jones[*], Shamith U. Payagala[*], Mattias Kruskopf[*,‡], Dean G. Jarrett[*], and Randolph E. Elmquist[*]

[*]National Institute of Standards and Technology, 100 Bureau Drive, Stop 8171, Gaithersburg, MD, 20899, USA
alireza.panna@nist.gov

[‡]University of Maryland, Joint Quantum Institute, College Park, MD, 20742, USA

***Abstract* — We describe recent scaling measurements to decade resistance levels based on both cryogenic and non-cryogenic current comparator bridges. National measurement institutes and the International Bureau of Weights and Measures derive traceability for the SI ohm using the quantum Hall effect, with the primary comparisons made at the resistance value $R_{K\text{-}90}/2$ = 12906.4035 Ω. Cryogenic current comparator scaling methods are preferred since large resistance ratios of 100 or more are achieved with unsurpassed uncertainty. The quantized Hall resistance standard based on graphene allow less complex and more cost-effective room-temperature resistance bridges to provide SI traceability.**

***Index Terms* — quantized Hall resistance, traceability, cryogenic current comparator, direct current comparator, standard resistor.**

I. Introduction

The National Institute of Standards and Technology (NIST) uses comparisons based on a bank of five 100 Ω standards and several cryogenic current comparator (CCC) bridges to initiate the traceability chain that ties four-terminal resistance standards to the quantized Hall resistance (QHR). The resistor bank has over 25 years of continuous measurement history at approximately half-yearly intervals based on a GaAs/AlGaAs heterostructure QHR device, prior to which traceability was derived from a bank of 1 Ω standards [1]. International consistency has been verified by resistor intercomparisons and on-site QHR comparisons [2]. One of the CCC bridges is a commercial system with binary windings and a wideband superconducting quantum interference device (SQUID) feedback [3]. This system allows exploration of many new measurement paths in scaling for four-terminal resistors, including precise direct comparisons between 1 kΩ and the QHR. Higher resistance levels from 100 kΩ to 10 MΩ are compared directly to the QHR using two-terminal CCC bridges with similar high stability obtained using DC SQUID feedback and a single voltage source [4].

Room-temperature direct current comparator (DCC) and dual-source bridge (DSB) systems can provide alternate resistance scaling paths without the need for cryogens and superconducting electronics. After the initial expense of the measurement system, these systems allow continuous operation with low maintenance cost. However, room-temperature bridge systems cannot achieve the sensitivity offered by high-turns CCC bridges, and an acceptable uncertainty level may be difficult to reach when scaling with a QHR standard. Higher current (or voltage) levels as required for good sensitivity and low type A uncertainty may introduce unwanted power dissipation in the standard resistor or quantized resistance breakdown [5] in the QHR device. Furthermore, since for many years liquid He and a high-field superconducting magnet have been required for operating the QHR device, most users have chosen to pursue CCC technology with the QHR standard.

Recently, the outlook for room-temperature QHR scaling has changed due to improvements in devices and instrumentation. The availability of graphene QHR standards with broad thermal and magnetic operating range [6, 7] allows their use with small superconducting magnets in tabletop cryocooler systems [8, 9]. The sensitivity of newer DCC bridges has also improved, providing QHR scaling at slightly higher device currents with uncertainty levels comparable to CCC bridges.

II. Measurement Design

In resistance metrology, the bridge ratio must be well known and stable to maintain reproducible uncertainties comparable to the stability of the standard resistors of the laboratory. If such ratios are maintained, room-temperature bridge systems could provide improved reliability at low cost based on year-round availability of the QHR standard. The precise ratios of a CCC system are used to calibrate DCC bridges and thus to improve the type B uncertainty; likewise, the digital voltage sources of the high-resistance DSB can be calibrated using Josephson array systems or resistance ratio voltage dividers so that the ratio uncertainty is known.

By direct comparison to CCC ratios, our measurements are intended as a step towards the determination of the long-term uncertainty of scaling using DCC bridges. Some ratios at NIST utilize a robust graphene QHR device in a table-top magnet system, as described in a separate report [9]. This ensures that the measurements have excellent long-term stability. Precision resistance standards are maintained at constant temperature for scaling and the type A uncertainty for each CCC ratio is derived as the standard deviation (SD) of a sequence of repeated measurement sets under very similar measurement conditions. The total time for each measurement set was of order 30 minutes, to compare type A results. The commercial binary-winding CCC (BCCC) was used to define most of the ratio values, with the number of windings giving optimum uncertainty for the current levels used. A two-terminal CCC

Elmquist, Randolph; Jarrett, Dean; Jones Jr., George; Kraft, Marlin; Kruskopf, Mattias; Panna, Alireza; Payagala, Shamith; Rigosi, Albert.
"Uncertainty of the Ohm Using Cryogenic and Non-Cryogenic Bridges."
Paper presented at Conference on Precision Electromagnetic Measurements (CPEM) 2018, Paris, France. July 8, 2018 - July 13, 2018.

(2TCCC) with winding ratio of 310-to-40 was used for measurements between the QHR standard and air-type 100 kΩ standard resistors.

III. Data and Results

DCC scaling results between the QHR standard and a 1 kΩ standard resistor are shown in Fig. 1. Here, the number of points averaged was varied inversely with the square of the applied voltage to obtain similar type A uncertainty for each point. The durations of the resulting data sets ranged from 110 min for the lowest measurement voltage (0.5 V) to 24 min for the highest (1.2 V), with the data from the first 10 min of each set discarded to allow the bridge nanovoltmeter balance to reach equilibrium. The error bars are SDs of the means, which underestimate the type A uncertainties due to short-term drift.

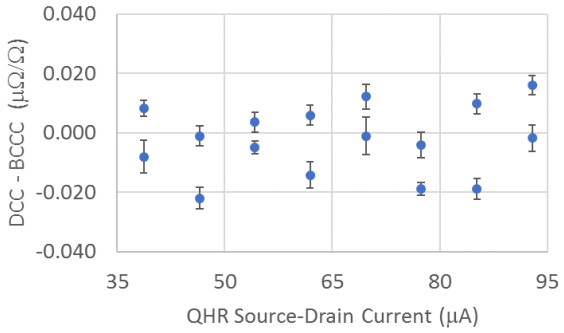


Figure 1. DCC bridge scaling from the QHR to a 1 kΩ standard using a range of source-drain currents. Results are normalized to the average results of BCCC scaling at 38.7 μA and 77.5 μA.

Preliminary standard uncertainty estimates (k = 1) of type A and type B influences in resistance scaling with the non-cryogenic DCC method are presented in Table 1. Some of these ratios were measured at the National Research Council (NRC) laboratory in Ottawa, Canada, in the initial calibration of the DCC bridge, and the values shown depend on those calibrations to help estimate long-term drift in the bridge ratios. Type B values represent combined estimates for all known influences.

Table 1. DCC standard uncertainty estimates

Resistance values (Ω)	Primary current (mA)	Secondary current (mA)	Meas. Time (min)	Type A (μΩ/Ω)	Type B (μΩ/Ω)
1 k, 100	1	10	20	0.003	0.007
10 k, 1 k	0.1	1	36	0.008	0.009
10 k, 1 k	0.32	3.2	36	0.004	0.009
R_K/2, 1 k	0.077	1	36	0.012	0.010
100 k, 10 k	0.01	0.1	32	0.096	0.062
100 k, 10 k	0.05	0.5	32	0.021	0.062

Similar results are given in Table 2, showing the uncertainty estimates (k = 1) of type A and type B influences in scaling ratios for the CCC bridges. Ratios with 100 kΩ use the 2TCCC.

Table 2. CCC standard uncertainty estimates

Resistance values (Ω)	Primary current (mA)	Secondary current (mA)	Meas. Time (min)	Type A (μΩ/Ω)	Type B (μΩ/Ω)
1 k, 100	1	10	20	0.0023	0.001
10 k, 100	0.1	10	20	0.0011	0.001
10 k, 1 k	0.1	1	20	0.0005	0.001
R_K/2, 100	0.0387	5	20	0.0008	0.001
R_K/2, 100	0.0775	10	20	0.0007	0.002
R_K/2, 1 k	0.0387	0.5	20	0.0005	0.001
R_K/2, 1 k	0.0775	1	20	0.0003	0.002
100 k, R_K/2	0.01	0.0775	30	0.011	0.025

IV. Conclusion

DCC scaling results between the QHR standard and decade resistance values from 100 Ω to 100 kΩ are compared to similar results from CCC measurements as a preliminary study of DCC suitability for traceability of the ohm based on the QHR. Relative uncertainties are estimated from these results, and will be more accurately determined for presentation at CPEM 2018.

References

[1] R. E. Elmquist, M. E. Cage, Y. Tang, A.-M. Jeffery, J. R. Kinard, Jr., R. F. Dziuba, N. M. Oldham, and E. R. Williams, "The Ampere and Electrical Standards," *J. Res., Natl. Inst. Stand. Technol.*, **106**, 65-103 (2001).

[2] F. Delahaye, T. J. Witt, R. E. Elmquist, and R. F. Dziuba, "Comparison of Quantum Hall Effect Resistance Standards of the NIST and the BIPM," *Metrologia* **37**, 173 – 176 (2000).

[3] D. Drung, M. Gotz, E. Pesel, H. J. Barthelmess, and C. Hinnrichs, *IEEE Trans. Instrum. Meas*. 62, 2820 (2013).

[4] F. L. Hernandez-Marquez, M. E. Bierzychudek, G. R. Jones Jr, and R. E. Elmquist, "Precision high-value resistance scaling with a two-terminal cryogenic current comparator*" Rev. Sci. Instrum.* 85 (4), 044701 (2014).

[5] B. Jeckelmann and B. Jeanneret, "The quantum Hall effect as an electrical resistance standard" *Rep. Prog. Phys.* 64 1603–1655 (2001).

[6] R. Ribeiro-Palau, R. F. Lafont, J. Brun-Picard, *et al*. "Quantum Hall resistance standard in graphene devices under relaxed experimental conditions" *Nature Nanotechnol.* 10, 965 (2015).

[7] Y. Yang, G. Cheng, P. Mende, *et al*. "Epitaxial graphene homogeneity and quantum Hall effect in millimeter-scale devices" Carbon 215, 229–236 (2017).

[8] T. J. B. M. Janssen, S. Rozhko, I. Antonov, A. Tzalenchuk, J. M. Williams, Z. Melhem, H. He, S. Lara-Avila, S. Kubatkin, and R. Yakimova, "Operation of a Graphene quantum Hall resistance standard in cryogen-free table-top system," *2D Mater.*, 2, 035015 (2015).

[9] A. F. Rigosi, et al., "A Table-Top Graphene Quantized Hall Standard" Submitted to CPEM 2018 (2018).

Elmquist, Randolph; Jarrett, Dean; Jones Jr., George; Kraft, Marlin; Kruskopf, Mattias; Panna, Alireza; Payagala, Shamith; Rigosi, Albert.
"Uncertainty of the Ohm Using Cryogenic and Non-Cryogenic Bridges."
Paper presented at Conference on Precision Electromagnetic Measurements (CPEM) 2018, Paris, France. July 8, 2018 - July 13, 2018.

Non-Absorbing, Point-of-Use, High-Power Laser Power Meter

Alexandra B. Artusio-Glimpse*, Ivan Ryger, Paul Williams, and John Lehman
National Institute of Standards and Technology, 325 Broadway, Boulder CO 80302, USA
**alexandra.artusio-glimpse@nist.gov*

Abstract: We have developed a compact, high-power laser power meter in the form of a folding mirror, precluding the need for beam splitters that considerably increase measurement uncertainty. Furthermore, our symmetric design inhibits responsivity to gravity.
OCIS codes: (120.3930) Metrological instrumentation; (120.3940) Metrology; (120.4290) Nondestructive testing

1. Introduction to radiation pressure power meters

Absolute laser power meters in the high-power regime ($\geq$100 W) traditionally require slow (tens of minutes) calorimetric measurements that exclude the beam under test from use downstream [1]. With the rise in industrial use of high-power laser processing, a need has arisen to develop high-accuracy point-of-use laser power meters that do not perturb the beam for its downstream use. This would facilitate desired quality control across platforms and over time for many commercially significant applications. Recently, Williams et al. [3] introduced an alternative primary standard laser power meter [4] that measures power from radiation pressure and traces its calibration of the optical watt to the kilogram. Derived from a commercial precision scale, this radiation pressure power meter (RPPM), while allowing use of the beam for processing throughout a measurement, is yet a relatively slow and bulky system filling a cube 30 cm on a side and having a response time of 5 s.

Smaller, faster, and more sensitive devices are needed to meet the measurement requirements of commercial systems. We, therefore, have improved upon and miniaturized the RPPM for convenient integration by incorporating microelectromechanical systems technology. To this end, we devised a capacitor-based force transducer that merged the optical elements (a high reflectivity mirror) and sensing elements into a compact, 4 cm on a side, package. In doing so, we increased the sensitivity by a factor of 100, decreased the response time of the detector by a factor of 50, and mitigated static sagging errors that arise when the device orientation changes with respect to gravity.

2. The "Smart Mirror" laser power meter

Our "Smart Mirror" capacitor-based compact RPPM sensor design is depicted in Fig. 1(a). The silicon micromachined force transducer consists of an Archimedean spiral planar spring supporting a circular plate with a high reflectivity mirror (reflectance >0.999 at 1070 nm) on one side and an electrode on the back side. An identical silicon spring is placed in close proximity to the first such that the two electrodes face each other forming a variable capacitor [5]. This dual spring configuration reduces the sensor response to sagging under gravity by an average factor of 22 dB over its single spring counterpart (see Fig. 1(b)). Mitigating sensor response to changes in orientation with respect to gravity is important for operating an embedded sensor at the end of a robotic arm, i.e. where a laser weld head may be placed – a constraint the bulk RPPM cannot meet.

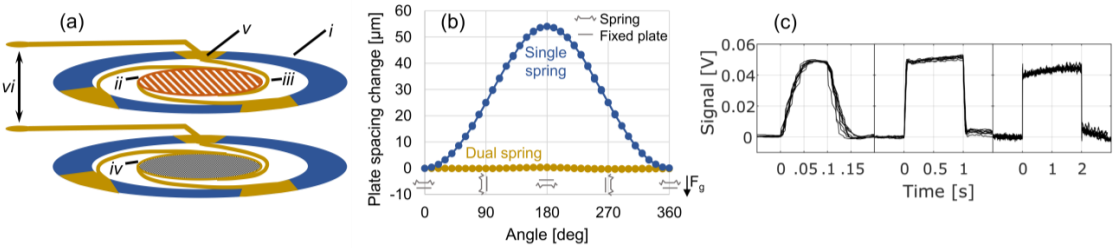


Fig. 1. (a) Depiction of dual spring sensor including mirror coated 10 mm diameter silicon disk *ii* attached to supporting silicon annulus *i* through 125 µm wide spring legs *iii* following an Archimedean spiral. Plate spacing is detected by measuring capacitance between two electrodes: *iv* and on the back side of *ii*. Electrical connection *vi* is provided through contact pads *v*. Springs are then clamped together in a mount with a plastic insulator providing a predefined spacing of 25 µm. (b) Measured change in static electrode spacing due to rotation of the sensor with respect to gravity (F_g) comparing a single spring and fixed plate electrode to the dual spring configuration. (c) Open-loop signal output of a single-spring prototype sensor upon 20 W CW 1070 nm laser injection at 45° incidence, square-wave modulated with (left to right) 0.1 s, 1 s, and 2 s periods. Each injected pulse was repeated six times and overlaid to show repeatability of sensor response to constant input. Standard deviation between repeated sensor response is 1 mV, or equal to the noise.

Artusio-Glimpse, Alexandra; Lehman, John; Ryger, Ivan; Williams, Paul.
"Non-Absorbing, Point-of-Use, High-Power Laser Power Meter."
Paper presented at OSA Imaging and Applied Optics Congress - Applied Industrial Optics Meeting, Orlando, FL, United States. June 25, 2018 - June 28, 2018.

The spring-based sensor is integrated with capacitive bridge electronics to convert the fractional change in sensor capacitance to an electrical voltage. The bridge output is then amplified and synchronously demodulated in a digital lock-in amplifier. This approach allows suppression of $1/f$ preamplifier noise and low frequency interference. A detailed review of the electronics and sensitivity optimization for this system is outlined in Ref. [6]. In prototype tests tracing the open-loop output of a single spring sensor, we measure noise levels below 1 W/√Hz, which agree with our expected value of 0.44 W/√Hz. We also observe noise-limited repeatability of the sensor response to a modulated CW laser impulse as seen in Fig. 1(c). Furthermore, we measure an 80 ms response time in open-loop. This response time is given directly by the natural resonance of the spring filtered by a 10 Hz noise bandwidth-limiting filter.

4. Calibrating lasers with embedded RPPMs

The "Smart Mirror" compact RPPM, as studied in these sensitivity and open-loop prototype tests, demonstrates improved measurement capability over its bulk predecessor. We find a low 1% single-to-noise ratio of the open-loop readout at 100 W laser power with a response time far faster than any other absolute power meter for high-power lasers. Like any folding mirror, this small, absolute power meter can be incorporated into laser optical systems for point-of-use laser calibration eliminating the need for beam splitters or pick-off techniques, which are sources of large uncertainty. The dual spring architecture for this sensor minimizes its response to static changes in the inertial reference frame. This adds flexibility to its use by allowing the sensor to be placed in additive manufacturing (AM) and laser welding systems where the laser head will move and rotate over the build time of a part. An absolute laser power meter of this type, embedded in laser processing systems, will facilitate process qualifications suitable for any so-equipped system. Currently, each individual AM or weld tool must be qualified for a particular process. RPPM technology offers a simple method for laser power calibration. By embedding these sensors into laser heads, we may look forward to systems where "calibrated lasers" can be exploited.

[1] Xiaoyu X. Li, Joshua A. Hadler, Christopher L. Cromer, John H. Lehman, and Marla L. Dowell, "High power laser calibrations at NIST," NIST Special Publication 250-77 (NIST, 2008).

[2] Mahesh Mani, Brandon Lane, Alkan Donmez, Shaw Feng, Shawn Moylan, and Ronnie Fesperman, "Measurement Science Needs for Real-time Control of Additive Manufacturing Powder Bed Fusion Processes," NIST Interagency/Internal Report (NISTIR) - 8036 (2015).

[3] Paul Williams, Joshua Hadler, Frank Maring, Robert Lee, Kyler Rogers, Brian Simonds, Matthew Spidell, Michelle Stephens, Ari Feldman, and John Lehman, "Portable, high-accuracy, non-absorbing laser power measurement at kilowatt levels by means of radiation pressure," Optics Express **25**(4), 4382-4392 (2017).

[4] Paul Williams, Joshua A. Hadler, Brian J. Simonds, and John H. Lehman, "Onsite multikilowatt laser power meter calibration using radiation pressure," Applied Optics **56**(34), 9596-9600 (2017).

[5] Alexandra Artusio-Glimpse, John Lehman, Michelle Stephens, Paul Williams, and Ivan Ryger, "Photon Momentum Sensor," US Patent Application 17-023US1 (2017).

[6] Ivan Ryger, Alexandra Artusio-Glimpse, Paul Williams, Nathan Tomlin, Michelle Stephens, Kyle Rogers, Matthew Spidell, and John Lehman, "Micromachined force balance for optical power measurement by radiation pressure sensing," Manuscript submitted for publication.

Artusio-Glimpse, Alexandra; Lehman, John; Ryger, Ivan; Williams, Paul.
"Non-Absorbing, Point-of-Use, High-Power Laser Power Meter."
Paper presented at OSA Imaging and Applied Optics Congress - Applied Industrial Optics Meeting, Orlando, FL, United States. June 25, 2018 - June 28, 2018.

Design of a cell-based refractometer with small end-effects

Patrick Egan and Jack Stone
National Institute of Standards and Technology, 100 Bureau Dr, Gaithersburg, MD 20899
Email: patrick.egan@nist.gov

***Abstract*—In cell-based laser refractometers, interferometer pathlength uncertainty introduced by deformation and stress in the windows through which the beams pass can be the chief factor limiting measurement accuracy. The fractional contribution of pathlength uncertainty to our recent determination of the Boltzmann constant was 9.8×10^{-6}, and more than two times larger than the next largest uncertainty component. We briefly describe the error, and propose a design in which cell window effects contribute less than 3×10^{-6} fractional error to the measurement of helium refractivity; performance that would be competitive with state-of-the-art primary thermometry and barometry.**

***Index Terms*—Interferometry, length metrology, refractive-index gas thermometry.**

I. Introduction

In a recent experiment [1], we used a cell-based laser refractometer called MIRE to determine the Boltzmann constant with 12.5×10^{-6} relative standard uncertainty, via measurements of helium refractivity [2] and the equation of state. Error arising from changes in window pathlength, induced by deformation and stress caused by the change in gas pressure, proved a debilitating uncertainty component, and produced a fractional error of 580×10^{-6} in our 25 cm-long cell. We attempted to cancel the error by measuring helium refractivity in cells of different lengths, and with almost identical material properties, window geometry, and location of beam transmission through all pairs of windows (i.e., cancellation of a common-mode error). After cancellation, the error contribution (9.8×10^{-6}, fractional) remained more than two times larger than the uncertainty in pressure and temperature measurements of the helium gas.

An additional drawback of the MIRE was the choice of crown glass for windows and spacer. This glass was chosen mainly for ease of manufacture, but its high and relatively unknown coefficient of thermal expansion meant that accurate refractivity measurements could only be carried out close to the temperature at which the cell length was measured (i.e., 20 °C). Thus, the uncertainty in thermodynamic temperature was the second largest uncertainty component in our determination of the Boltzmann constant, and our original choice of glass has precluded MIRE from performing as an accurate refractive-index gas thermometer over a broader range.

II. Designs for small end-effects

As discussed by Shelton [3], the change in pathlength through a window exposed to pressure $d_w\cdot p=(n_i-1)\cdot(w_f-w_i)+w_f\cdot(n_f-n_i)$ consists of two terms: the first arising from a change in the geometric thickness of the window from initial w_i to final w_f, and the second term from a change in the refractive index of the glass from initial n_i to final n_f. For a pressure increase inside the cell, the first term is negative and the second term is positive (in general) and smaller than the first term. In other words, an increase in pressure inside the cell decreases the optical pathlength through the window. In our case, we estimate the change in geometric thickness from finite-element analysis, and the change in refractive index from photoelastic coefficients and applied stress [4], [5], where the estimate of stress also comes from finite-element analysis.

MIRE was designed around a triple-cell of Fig. 1(a), as this seemed an expedient way of passing two beams through a center cell at the same gas pressure and temperature, and two beams through the outer cells at vacuum. However, this arrangement meant the diameter of the bores were 19 mm, and $d_w=-23.75$ fm/Pa; furthermore, the design had beams passing near the edges of the bores where the sensitivity of window pathlength to changes in beam location was $\frac{\mathrm{d}}{\mathrm{d}x}d_w=3.0\,\frac{\mathrm{fm/Pa}}{\mathrm{mm}}$. This sensitivity coefficient is important when considering how well the error can be corrected via measurements of refractivity with cells of different lengths. We note two things: (1) our previous effort with alignment achieved four

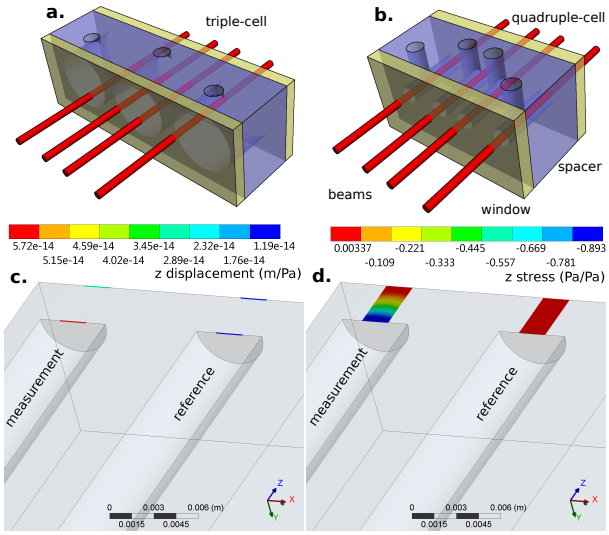


Fig. 1. (a) Sketch of the original MIRE 1.8 cm triple-cell. (b) A quadruple-cell. (c) Distortion and (d) stress on a fused-silica window at the point of transmission for one pair of measurement and reference beams [i.e., a quarter-section of the geometry shown in (b)].

Egan, Patrick; Stone Jr., Jack.
"Design of a cell-based refractometer with small end-effects."
Paper presented at Conference on Precision Electromagnetic Measurements (CPEM) 2018, Paris, France. July 8, 2018 - July 13, 2018.

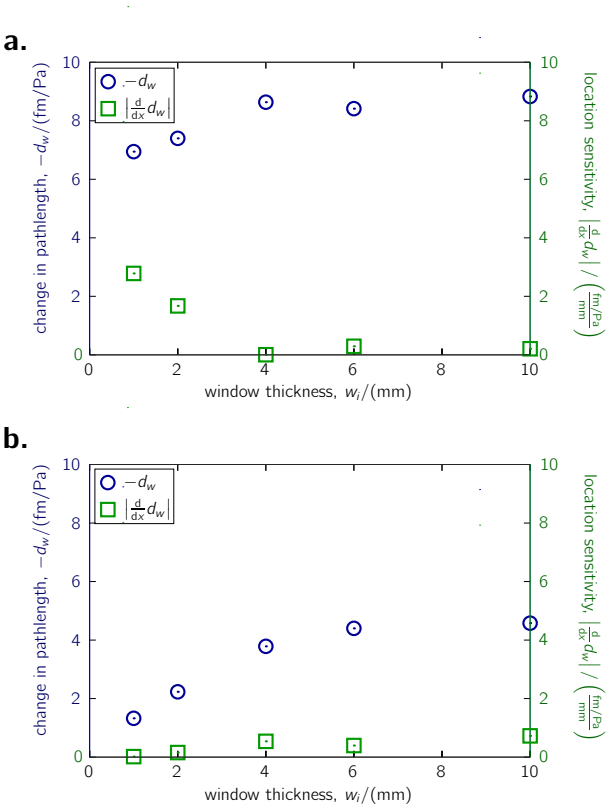


Fig. 2. Change in optical pathlength through (a) fused-silica and (b) sapphire windows. Sensitivity of the window pathlength to change in beam location shown on second y-axis.

beams parallel to $\pm 85\,\mu$rad, and thus beam location through long and short cells differed by as much as 74 µm on one window; and (2) a variable-length cell would use the same windows for long and short cell lengths, and could achieve a very effective cancellation, but here we wish to avoid the complication of motion and external metrology, and consider cancellation with two or more cells of very different lengths.

There are two obvious ways to reduce the larger $w_f - w_i$ term: (1) reduce the area upon which the pressure acts, and (2) increase the elastic modulus of the window material. A quadruple-cell design can reduce the exposed area and also center the beam location in the bore. An all fused-silica design is shown in Fig. 1(b–d), with one 5 mm diameter hole for each of the four 2 mm diameter beams, passing through 4 mm thick windows. The results of finite-element and photoelastic calculations are shown in Fig. 2(a), where the change in window pathlength is plotted as a function of window thickness. A window thickness of between 4 mm and 6 mm would be optimum in ensuring $d_w \approx -8$ fm/Pa, and this thickness range also exhibits low sensitivity to changes in beam location $\frac{\mathrm{d}}{\mathrm{d}x}d_w < 0.5\,\frac{\text{fm/Pa}}{\text{mm}}$. In an all fused-silica quadruple-cell refractometer, we thus expect a factor of 3 or more reduction in the uncertainty of the window pathlength change compared to our previous triple-cell design. A few more points are worth noting: (1) an additional benefit of the small diameter bore is the possibility to reach pressures above 1 MPa where errors in refractivity measurements due to uncertainties in cell length and interferometer phase are proportionally smaller; (2) the thermal expansion of fused-silica is low and well-known enough that uncertainty in the length of the cell would not be a dominant component if used as a refractive-index gas thermometer in the range (293 ± 100) K; and (3) fused-silica is available in precision-bore tubing, which could simplify cell manufacture and allow cell lengths of 50 cm and longer, more than twice as long as our previous design (i.e., a further reduction of 2 in the window pathlength uncertainty contribution).

The alternative way to reduce $w_f - w_i$ would be to use a quadruple-cell design with sapphire windows and a spacer of low thermal expansion glass-ceramic. In Fig. 2(b) we plot pathlength change and sensitivity to beam location for sapphire windows, where the optical axis is parallel to the c-axis of the crystal, and geometry is the same as Fig. 1. For a window thickness of 2 mm, the change in pathlength is about four times smaller than fused-silica, and the sensitivity to beam location is similarly low. However, the large difference in thermal expansion between window and spacer leads to a temperature-dependent change in pathlength of $d_w = 9.6$ pm/mK for $w_i = 2$ mm, and 19.0 pm/mK for $w_i = 4$ mm. This effect would likely be the limiting factor toward achieving sub-3×10^{-6} relative uncertainty in the measurement of helium refractivity. In our past experiments, a fill from vacuum to 367 kPa increased cell temperature by 40 mK, and after 4 h cell temperature was still almost 10 mK above its initial temperature (i.e., the point at which the initial interferometer phase was measured). It thus seems likely that during gas measurements the change in interferometer phase would be uncertain at the 100 pm-level.

III. Conclusion

We have proposed a design for a laser refractometer that employs a quadruple-cell. We expect an all fused-silica cell with 5 mm diameter bore-holes should reduce uncertainty due to changes in window pathlength by a factor of 3 compared to our previous design [1]. The design should make possible realizations of either the pascal or kelvin to within 3×10^{-6} relative standard uncertainty, via measurement of helium refractivity and the equation of state [2]. An alternative design with sapphire windows and a low thermal expansion spacer would have lower pressure-induced changes in pathlength, but mismatch in thermal expansion and the resulting temperature-induced changes in pathlength are a major concern.

References

[1] P. F. Egan, J. A. Stone, J. E. Ricker, J. H. Hendricks, and G. F. Strouse, "Cell-based refractometer for pascal realization," *Optics Letters*, vol. 42, no. 15, pp. 3945–3948, 2017.

[2] M. Puchalski, K. Piszczatowski, J. Komasa, B. Jeziorski, and K. Szalewicz, "Theoretical determination of the polarizability dispersion and the refractive index of helium," *Physical Review A*, vol. 93, 032515, 2016.

[3] D. P. Shelton, "Lens induced by stress in optical windows for high-pressure cells," *Review of Scientific Instruments* vol. 63, no. 8, pp. 3978–3982, (1992).

[4] H. Bach and N. Neuroth, eds., *The Properties of Optical Glass*, Springer–Verlag, 1998, ch. 2.

[5] B. D. Guenther, *Modern Optics*, 2nd ed. Oxford University Press, 2015, ch. 13 & 14.

Egan, Patrick; Stone Jr., Jack.
"Design of a cell-based refractometer with small end-effects."
Paper presented at Conference on Precision Electromagnetic Measurements (CPEM) 2018, Paris, France. July 8, 2018 - July 13, 2018.

Glassy-Electret Random Access Memory – A naturally Nanoscale Memory Concept

Vasileia Georgiou[1,2], Jason P. Campbell[1], Pragya R. Shrestha[1,3], Dimitris E. Ioannou[2], and Kin P. Cheung[1]*
1.National Institute of Standards and Technology, Gaithersburg, USA
2.Department of Electrical and Computer Engineering, George Mason University, United States.
3.Theiss Research, La Jolla, California, United States. * E-mail: kin.cheung@nist.gov

The self-heating effect (SHE) is a growing problem for decananometer CMOS and beyond with substantial efforts dedicated to mitigation. Here, we present a new memory concept which instead requires SHE exacerbation. As such, this memory concept is naturally suitable for extreme scaling. Preliminary result of this memory concept is demonstrated with an external heater as the SHE surrogates.

This memory is based on glassy-electrets. Glassy-electrets are poled polymers with glass-transition temperature, T_g, well above the operation temperature. When used as a ferroelectric memory device, the glassy-electret polymer is rapidly poled (programmed) at temperatures well above T_g (a fully molten polymer switching time < 1 ps). Upon cooling, the polarization (memory state) is "frozen" and has long retention and disturb immunity.

The Glassy-Electret Random Access Memory (GeRAM) shown in fig. 1 is a ferroelectric field-effect transistor (FeFET) memory with the ferroelectric gate dielectric replaced by glassy-electret. This key difference eliminates known FeFET problems [1, 2] such as depolarization, imprint, disturbs, etc., and adds multi-state capability and low voltage operation.

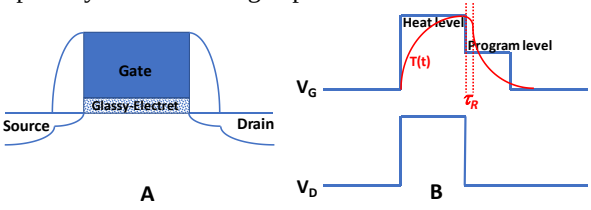


Fig. 1, A: GeRAM with Glassy-Electret as gate dielectric of a MOSFET; B: Programming pulse waveform. T(t): temperature profile; τ_R: dipole rotation time.

GeRAM programming involves the application of an electric field across the gate dielectric (polarization) while the device is held at a higher temperature (above Tg). This high temperature is provided by the SHE of the transistor. Instead of engineering the transistor to minimize SHE, one would enhance SHE by minimizing heat loss, very similar to the efforts being made in Phase-Change Memory (PCRAM) development [3]. SHE temperature increases (ΔT) of >100 C have been reported [4, 5] in the channel. With the appropriate thermal barrier engineering, it is entirely reasonable to expect ΔT extending up to 300 C.

The required ΔT is a function of the chosen polymer material. The material should ideally support <100 ps program time and >10 years retention time. This corresponds to a dipole rotation rate that varies ~19 orders of magnitude between the program and operation temperatures. Fortunately, there are existing polymers which almost meet these specifications. Fig. 2 shows the peak response frequency extracted from dielectric spectroscopy for four polar polymers plotted against temperature.

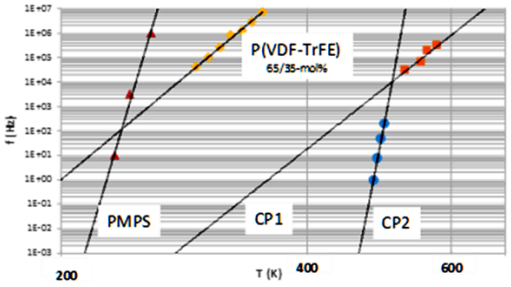


Fig. 2 Peak respond frequency as a function of temperature for 4 polar polymers. CP1 (2,2-bis(3-aminophenyl)hexafluoropropane +2,2-bis[4-(4-aminophenoxy)phenyl] hexafluoropropane) was measured in this work while the other three are extracted from literature [6-8].

The slopes for PMPS and CP2 are 5.1 C/dec and 6.5 C/dec respectively, meaning that PMPS needs ΔT=96.9 C to span 19 orders of magnitude and CP2 needs ΔT=123.5 C. The extrapolation over 19 orders of magnitude using limited range data is of course quite risky. However, at least one report exists showing the viscosity of chalcogenide glass varies smoothly as a function of temperature over 17 orders of magnitude [9]. Indeed, the trend is not quite a straight line in the log-log scale so that the straight-line extrapolation is on the pessimistic side.

T_g for CP2 is ~200 C and 100 ps program time requires ~265 C. Thus, GeRAM utilzing CP2 as the gate dielectric needs ΔT to be ~240 C. Therefore, at an operation temperature of 70 C, the retention time exceeds 10^{20}s.

Assuming the minimum heat pulse duration equals the program time, one can estimate the energy required for programming. Compared to PCRAM, where the

Campbell, Jason; Cheung, Kin; Georgiou, Vasileia; Ioannou, D.; Shrestha, Pragya.
"Glassy-Electret Random Access Memory - A naturally Nanoscale Memory Concept."
Paper presented at The 2018 International Symposium on VLSI Technology, Systems and Applications, Hsinchu, Taiwan Province of China.
April 16, 2018 - April 20, 2018.

temperature is higher and the duration is much longer, one can see that GeRAM is intrinsically more energy efficient, by as much as 3 orders of magnitude. The steep slope of the frequency vs. temperature curve also eliminates program disturb. Decreasing the energy by half decreases the ΔT by half, leading to $>10^{18}$ times slower response.

As a proof of concept, we fabricated a GeRAM using 2 x 2 μm^2 polysilicon junction-less thin-film transistors (Fig. 3). Since this is not a high-performance nanoscale MOSFET, it cannot provide adequate SHE. Therefore, a hot-chuck provides the needed temperature gradient, but the speed is much slower (~60 s instead of 100 ps).

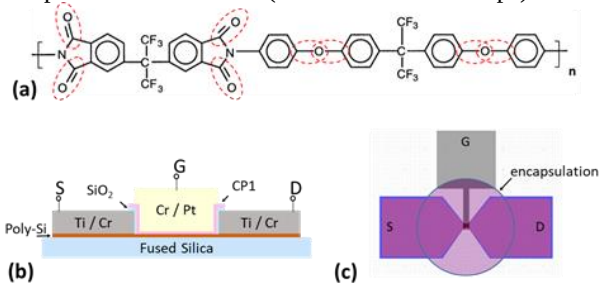


Fig. 3 (a) CP1 polymer; (b) cross section of the GeRAM fabricated; (c) top view of the GeRAM fabricated.

The junction-less TFT has, admittedly poor, gate modulation < one order of magnitude. We did not optimize the TFT as it was only needed to sense the polarization change. This is an obvious area for improvement in future fabrication cycles. Instead of the optimal CP2, the TFT's were fabricated with CP1 polymer gate dielectrics (13 nm). CP1 was chosen due to its commercially availability.

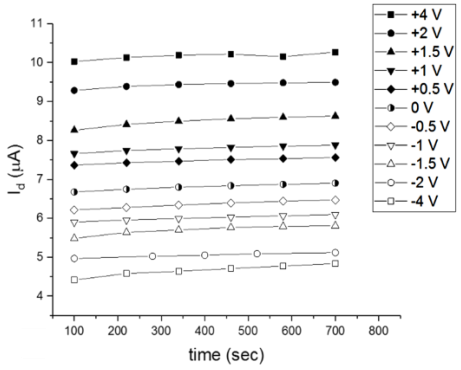


Fig. 4 Multi-states programming of the fabricated GeRAM. Program was performed at 275 C. Measurement was done at room temperature.

Fig. 4 shows programming results for various program voltages. Programming was done at 275 C and sense measurements were done at room temperature. In between each programming, the device was erased by keeping it at 275 C for 60 s with no applied voltage. 275 C is 15 C above T_g for CP1, so 60 s is sufficient to randomize the dipoles.

When the dipoles are free to rotate, programming is accomplished at any applied field and the net dipole moment will satisfy the applied field. This enables multi-states programming, as shown in fig. 4. More importantly, the programed states are immune to room temperature application of 4 V gate voltage (both positive or negative), for 10 minutes. This illustrates the robustness of GeRAM.

The remaining question lies in the endurance of the cell. If each program step heats the polymer to near melting (way above T_g) to achieve high program speed, how many times can it be programmed? Naturally, it depends on the material properties of the polymer. Many polymers remain stable in a molten state for days or longer in the absence of oxygen (or an oxidizing agent). It is expected the GeRAM will be fully encapsulated eliminating oxidation concerns. Realization of 100 ps program times, translates one day to ~10^{15} cycles.

Since the polymer is nearly melted during each cycle, full relaxation is insured for each cycle. Thus, imprint problems should not be a concern for GeRAM.

How far can GeRAM scale? Polar polymers with dipole spacing down to 0.6 nm exists. Thus a 3 x 3 x 3 nm^3 gate dielectric can contain up to 125 dipoles – sufficient for reliable operation.

Finally, the question of processing must be addressed. The need to withstand 400 C for 30 min. may prohibit lower T_g polymers even if there are energy advantages. CP1 was held at 310 C in air for nearly two hours with no problems. Thus, we expect it to be stable at 400 C for 30 min after full encapsulation. An advantage of polar polymer is the flexibility of tuning the property through synthesis – once the desired property is known.

In summary, a new memory concept is proposed. It is well suited for extreme scaling and leverages the SHE of decananometer transistors. It has the potential for very high-speed operation, and for long retention and high endurance. The potential for high-density is also very promising.

[1] Ma, T. P. and J.-P. Han (2002). IEEE EDL **23**(7): 386.
[2] Pan, X. and T. P. Ma (2011). APL **99**(1): 013505-013505-3.
[3] H. Hayat, K. Kohary, and D. Wright, Nanotechnology 28 (2017) 035202.
[4] S. E. Liu, J. S. Wang et. al., IRPS 2014, 4A.4.1.
[5] S. H. Shin, S. H. Kim et. al., IEDM2016, 15.7.4.
[6] V. Bharti, Q. Zhang, Phys. Rev. B 2001, 63, 1.
[7] K. L. Ngai, Eur. Phys. J. E 2002, 225.
[8] J. D. Jacobs, M. J. Arlen, et. al., Polymer (Guildf). 2010, 51, 3139.
[9] J-L. Adam, X. Zhang, (2014). "Chalcogenide Glasses: Preparation, Properties and Applications." Elsevier Science. p. 94. ISBN 978-0-85709-356-1.

Campbell, Jason; Cheung, Kin; Georgiou, Vasileia; Ioannou, D.; Shrestha, Pragya.
"Glassy-Electret Random Access Memory - A naturally Nanoscale Memory Concept."
Paper presented at The 2018 International Symposium on VLSI Technology, Systems and Applications, Hsinchu, Taiwan Province of China.
April 16, 2018 - April 20, 2018.

Imaging and Applied Optics 2018, The Optical Society of America, Paper AM3A.4, Orlando, Florida, USA
25–28 June 2018, https://doi.org/10.1364/AIO.2018.AM3A.4

Uniform Angular Illumination in Optical Microscopes

Ravi Kiran Attota*, Emil Agocs
Engineering Physics Davison, PML, NIST, Gaithersburg, MD 20899, USA
Author e-mail address: Ravikiran.attota@nist.gov

Abstract: Angular illumination asymmetry (ANILAS) at the sample plane depends on illumination wavelength, objective type and the location of aperture stop. To extract consistent and accurate quantitative values, all the three parameters must be aligned.
OCIS codes: (170.2945) Illumination design; (170.0110) Imaging systems

1. Introduction

Non-uniform angular illumination aberration across the field-of-view is one of the often-overlooked optical aberrations that has significant effect on quantitative imaging, despite having the uniform spatial intensity with Koehler illumination design. The non-uniform angular illumination results in a loss of imaging precision and accuracy leading to less reliable quantitative measurements. Most of the commercially available optical microscopes are generally well designed and aligned. However, these microscopes have potential to readily lose their optimal illumination by misalignment of either the illumination source or the aperture diaphragm. In this paper, we mainly study the effect of misalignment of the aperture diaphragm.

In the referenced work [2-3], we proposed a simple and fast method to measure the angular illumination asymmetry (ANILAS) at the sample plane. By iteratively evaluating the ANILAS maps with careful alignment of the optical elements, it is possible to achieve the lowest distortion in the angular illumination for a given objective. However, as we reveal here that the set of optimized conditions is only good for that particular objective and illumination used. There is a good chance that simply selecting a different objective or illumination wavelength could lead to a sub-optimal illumination conditions, even if all the other conditions remain the same. Here we demonstrate that nearly every objective requires its own optimal alignment condition that helps to enhance optical imaging precision and hence obtain consistent quantitative values.

2. Results

A commercially available, research-grade optical microscope was used for the experiments [3] with different magnification and numerical aperture (NA) objective. We used three LED illumination sources (along with band-pass filters) to produce a narrow-band illumination centered around 405 nm ± 5 nm, 520 nm ± 5 nm and 633 nm ± 2 nm. ANILAS maps were evaluated using the dense grating method as described in an earlier publication [3] using an array of trenches in SiO_2 with a nominal width of 100 nm having a pitch of 1000 nm over a Si substrate as the grating target.

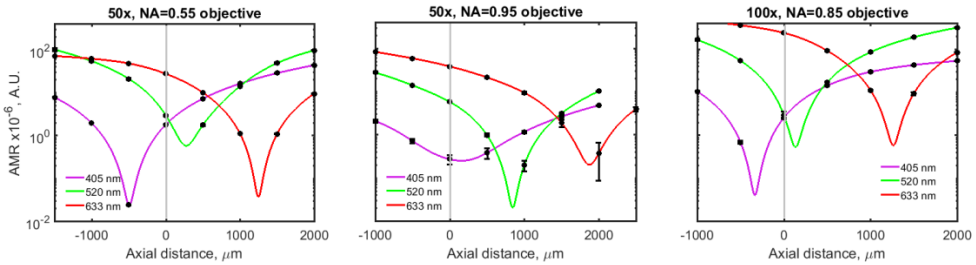


Fig. 1 Plots of ANILAS magnitude ranges (AMR) with the aperture diaphragm axial location for the three objectives studied (a) 50x, NA=0.55, (b) 50x, NA=0.95 and (c) 100x, NA=0.85. The wavelengths of illumination are shown in the legend. Data points (filled circles) were fitted with cubic spline curves. Grey vertical line represents the axial location of the original aperture diaphragm.

The axial location where the minimum AMR occurs in Fig. 1 represents the best axial location of the aperture diaphragm that results in the lowest illumination distortion for that objective, i.e. nearly uniform and symmetric

Agocs, Emil; Attota, Ravikiran.
"Uniform Angular Illumination in Optical Microscopes."
Paper presented at Applied Industrial Optics, Orlando, FL, United States. June 25, 2018 - June 28, 2018.

angular illumination over the entire FOV. From this point any axial deviation of the aperture diaphragm on either side results in increased degree of illumination non-uniformity in the FOV.

Based on this explanation, a few important observations can be made from Fig. 1. The first one is that the axial aperture diaphragm location where the lowest distorted illumination occurs varies with the objective type used even if all the other conditions are the same. In other words, every objective has its own optimum axial location for its lowest illumination distortions. This implies that the one fixed axial aperture diaphragm location provided in most of the optical microscopes cannot physically match with the optimum axial illumination locations for all the objectives.

A second observation is that the optimum axial location not only depends on the type of objective, but also on the illumination wavelength. For the objectives tested, the minimum AMR location moves away from the field aperture (toward the positive axial distance in the convention used here) with increasing illumination wavelength.

Minimum AMR distance data shown in Fig. 1 when presented as a function of the wavelengths as shown in Fig. 2 reveals some additional useful information. If the original aperture diaphragm axial location cannot be moved axially for illumination optimization (which is the case for most of the optical microscopes), this figure enables to determine the wavelength at which a given objective produces the lowest illumination distortions. From Fig.2 we can determine that the objectives 1, 4 and 6 have the best illumination if used with approximately 480 nm, 365 nm and 500 nm illumination wavelengths, respectively. Conversely, if the original aperture diaphragm location can be moved axially for illumination optimization, this figure enables to determine the axial location of the aperture diaphragm where a given objective produces the lowest distorted illumination. For example, if the objective number 4 is desired to be used in combination with an illumination wavelength of 600 nm, then we can determine that the aperture stop axial location of approximately 1.5 mm provides the best illumination condition (red dashed arrows in Fig. 4). It also shows that for the visible spectrum and the objectives used in the current optical microscope, it requires an aperture diaphragm axial alignment range of over 3 mm to achieve the best illumination condition.

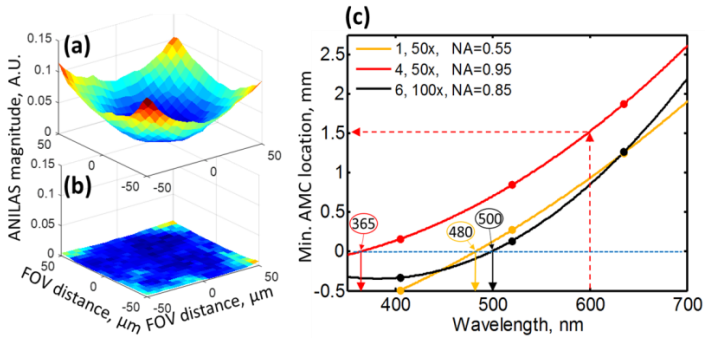


Fig. 2 Typical ANILAS maps for (a) poor, and (b) good angular illuminations. (c) A plot of the minimum AMR axial distances as a function of the illumination wavelengths for the three objectives. Horizontal blue dashed line shows the axial location of the original aperture diaphragm. The solid arrows pointing down from the blue dashed line represent the wavelengths at which the minimum AMR coincides with the original aperture axial location. The objective number, magnification and NA are shown in the legend, in that order.

3. Conclusion

ANILAS maps provide a convenient way to measure and visualize the quality of illumination at the sample plane. The axial location of the aperture stop corresponding to the lowest distorted illumination depends upon the objective lens type and the illumination wavelength, with a spread of about 2.36 mm in the current study conditions. Optical images for precision, quantitative imaging, and for metrology applications, this illumination aberration must be minimized by aligning the aperture diaphragm. Hence, we propose to microscope manufacturers that they provide sufficient axial alignment capability for aperture stops (in addition to lateral alignment capability).

4. References

[1] This summary is based on the following article: E. Agocs and R. K. Attota, "Enhancing Optical Microscopy Illumination for Quantitative Imaging,", Scientific Reports, under review.
[2] R. Attota and R. Silver, "Optical microscope angular illumination analysis," Opt. Express 20, 6693-6702 (2012).
[3] R. K. Attota, "Step beyond Kohler illumination analysis for far-field quantitative imaging: angular illumination asymmetry (ANILAS) maps," Opt Express 24, 22616-22627 (2016).

Agocs, Emil; Attota, Ravikiran.
"Uniform Angular Illumination in Optical Microscopes."
Paper presented at Applied Industrial Optics, Orlando, FL, United States. June 25, 2018 - June 28, 2018.

Progress Towards Accurate Monitoring of Flue Gas Emissions

Aaron Johnson[1], Iosif Shinder, Michael Moldover, Joey Boyd, James Filla
Sensor Science Division, NIST, Gaithersburg, MD 20899

Abstract

The amounts of CO_2 and other pollutants emitted by a coal-fired power plant are measured using a continuous emissions monitoring system (CEMS) permanently installed in the exhaust smokestack. The pollutant flux is the product of the pollutant's concentration and the flow in the stack. The concentration measurements are traceable to certified reference standards; however, the complex velocity fields (i.e., swirling flow with a skewed velocity profile) in stacks make accurate flow measurements difficult. Therefore, the CEMS flow monitor, which commonly consists of a single-path ultrasonic flow meter, must be calibrated at least once a year by a procedure called a relative accuracy test audit (RATA). This calibration is generally performed using a differential pressure probe called an S-type pitot probe. However, S-probes are not accurate when used in complex velocity fields. NIST developed a 1/10th scale-model smokestack simulator (SMSS) to quantify the uncertainty of diverse stack flow measurement techniques. The SMSS generates complex, stack-like flows, but with an expanded uncertainty (95 % confidence level) in the average velocity of 0.7 %. Using the SMSS, we assessed S-probe-based RATAs and both single and two-crossing-path (X-pattern) CEMS ultrasonic flow monitors. Remarkably, the X-pattern ultrasonic CEMS deviated by only 0.5 % from the NIST's flow standard. In contrast, a single-path ultrasonic CEMS deviated from NIST standards by 14 % to 17 % in a highly distorted flow. Deviations for the S-probe RATAs ranged from 5 % to 6 %.

1 Introduction

The combustion gases from coal-fired power plants (CFPPs) are exhausted into vertical, large diameter (D_{stack} > 5 m) smokestacks. These stacks disperse combustion gases high in the atmosphere to minimize pollution at the ground level. To quantify the amount of pollutants released into the atmosphere, the total flow in the stack must be accurately measured. However, stacks, are not designed to facilitate accurate flow measurements. The network of elbows, reducers, fans, *etc*. upstream of the stack inlet generate complex velocity fields that make accurate stack flow measurements difficult. Additional difficulties occur because the flue gas from CFPPs is hot (around 50 °C), saturated with water vapor, and depleted in oxygen (asphyxiating). Despite the difficulties, pollutant emissions are measured by Continuous Emissions Monitoring Systems (CEMS) installed in these smokestacks.

CEMS systems measure both the concentration and the total flow in the stack. The product of the concentration and the total flow is the instantaneous rate of pollutant efflux. Since 1992 NIST[2] has supported accurate CEMS concentration measurements by providing certified gas mixtures called NIST Traceable Reference Materials (NTRMs) to the stack testing community [1]. These NTRMs are traceable to NIST primary standards at specified uncertainty levels, and are often used as standards to verify the accuracy of stack concentration measurements [2 - 5]. Before the present work, NIST

1 Corresponding Author: aaron.johnson@nist.gov
2 The National Institute of Standards and Technology (NIST) is the U.S. National Metrology Institute.

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

was not involved in establishing standards for accurate stack flow measurements. Yet, the flow measurement is arguably more significant than the concentration measurement. An erroneous measurement of the concentration of a pollutant affects the mass calculation of that one pollutant, while an erroneous flow measurement is multiplied by all the concentration measurements and therefore leads to erroneous values of *all* emitted pollutants. Furthermore, *all* of the pollutant values are biased in the same direction as the erroneous flow measurement.

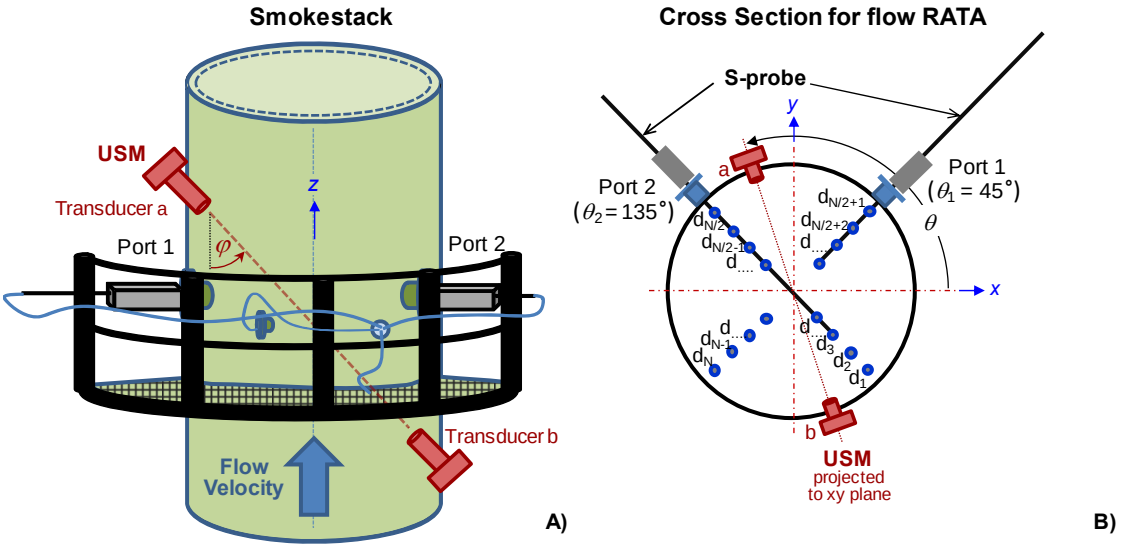


Figure 1. Sketch of catwalk surrounding a smokestack with installed apparatus for measuring the flow in the smokestack. A) single-path Ultrasonic Flow Monitor (USM) installed at path angle φ, and two orthogonal ports (Port 1 and Port 2) used for the flow RATA, and B) stack cross section showing locations $d_1 \ldots d_N$ where axial flow velocities are measured by traversing an S-probe through ports 1 and 2. In this figure θ denotes the installation angle of the two ports and of the USM.

In CFPPs the most widely used CEMS flow monitor is a single-path ultrasonic flow meter (USM) [6]. Figure 1 depicts a single-path USM installed in a stack at a path angle φ and at an installation angle of θ. The axial flow velocity is determined using the expression

$$\bar{V}_{\mathrm{USM,f}} = k\,\bar{V}_{\mathrm{USM,p}} / \cos\varphi\,, \qquad (1)$$

where k is the experimental (or theoretical) calibration factor; and the bar above the velocity "$\bar{\ }$" denotes that the velocity is the average value over the path. As such, $\bar{V}_{\mathrm{USM,p}}$ is the path-averaged velocity directed along the acoustic path indicated by the dashed line in Fig. 1A[3]. This velocity is determined by measuring the time it takes ultrasonic pulses to travel the distance between transducers a and b with and against the flow. If the transit times are accurately measured and the path length (L_p) is known, $\bar{V}_{\mathrm{USM,p}}$ can be accurately determined. However, the velocity that is essential for accurate emissions measurements is $\bar{V}_{\mathrm{USM,f}}$, which is directed along the stack axis.

3 Throughout this document USM quantities with the subscript "p" indicate the value of that quantity along the acoustic path (L_p is the path length, a_p is average sound speed along the path, etc.). The subscript "f' is used to denote the USM velocity directed along the stack axis (*i.e.*, the flow velocity).

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Unfortunately, $\overline{V}_{USM,f}$ cannot in general be accurately determined by a single path USM. For stack flows, the uncertainty of an uncalibrated single path USM typically ranges from 5 % to 20 %. The cause of this error is two-fold. First, the geometric factor 1/cosφ in Eq. (1) is intended to convert the velocity directed along the acoustic path to the velocity along the stack axis. However, if the cross-flow velocity (V_c), commonly called swirl, is not zero, then the resulting velocity is not directed along the stack axis and is therefore not indicative of the flow velocity. Second, the USM only provides a path-averaged velocity instead of the area weighted velocity required to determine the average flow. The difference between the path and area weighted velocity can be substantial when the axial velocity profile is not non-uniform. We herein refer to this error as a profile effect. Because stack flows have complex velocity fields with non-zero cross-flows and asymmetric velocity profiles, the swirl and profile effect errors of single path USMs are generally too large to provide sufficient accuracy for emission measurements. To improve CEMS flow measurement accuracy, the USMs are calibrated at least yearly by an EPA protocol called a relative accuracy test audit (RATA). This RATA generally uses an S-type pitot probe to measure the axial velocity at prescribed locations along the cross section of a stack [7 - 9]. Figure 1B illustrates how an S-type pitot probe is inserted into the stack access ports and traversed along the 2 orthogonal chords to the prescribed measurement locations. The velocity measurements made at points d_1 . . . d_N are area-weighted to determine the average flow velocity, which is subsequently used to determine the calibration factor k in Eq. (1).

Although current regulations require CEMS flow measurements to be calibrated against the flow RATA, this correction does not necessarily improve the flow measurement accuracy. The accuracy of the S-probe, like the USM, degrades in the complex velocity fields inherent to stack flows. The S-probe's accuracy degrades because this probe is only designed to measure 2 components of the velocity vector. Researchers have shown that when the third component of velocity (*i.e.*, the velocity along the pitch angle) differs between calibration and application, large errors can result [10, 11]. In stack applications, S-probes are calibrated in a wind tunnel at zero pitch angle, but are used in stacks where a typical pitch angle ranges from $= -10\,^\circ \leq \alpha \leq 10\,^\circ$. More extreme pitch angles are possible, depending on the piping configuration and flow conditions immediately upstream of the stack inlet. Shinder showed that when a calibrated S-probe is used at large negative pitch angles (*e.g.*, $\alpha = -10\,^\circ$) errors can exceed 6 % [10]. Moreover, many flow RATAs are performed using uncalibrated S-probes, which can introduce errors of 10 % or more [12, 13].

Flow RATA errors are transferred to CEMS flow measurements and can result in unknown biases in stack flow measurements. In this work, we assess the accuracy of an X-pattern USM. The single-path USM in Fig. 1 can be converted to the X-pattern design by installing a second USM at the same path angle ($\varphi_2 = \varphi_1$) and spaced 180 ° apart from the first USM ($\theta_2 = \theta_1 + 180\,^\circ$). The X-pattern is shown in Fig. 6. The X-pattern flow velocity is defined as the arithmetic average of the flow velocities of the 2 single-path USMs. To avoid introducing a potential bias from the S-probe flow RATA we did not flow calibrate the X-pattern or the single path USM in this research (*i.e.*, $k = 1$ in Eq. (1)). We assessed the absolute accuracy of these uncalibrated CEMS in smokestack-like flow conditions using a facility called the Scale-Model Smokestack Simulator (SMSS) [14]. Remarkably, we found average velocity measured by the X-pattern CEMS was within 1 % of the NIST traceable reference velocity in the SMSS facility. In contrast, the accuracy of a single-path USM ranged from 5 % to 17 % depending

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

on its θ orientation. The accuracy of the S-probe RATA ranged from 5 % to 6 % depending on the number of traverse points, flow conditions, and θ orientation.

This manuscript documents the performance of CEMS and RATA flow measurements in NIST's SMSS facility. Section 2 describes the SMSS, which generates smokestack-like flow conditions in a 1.2 m diameter Test Section, but measures the flow with an expanded uncertainty (corresponding to a 95 % confidence level) of 0.7 % using a reference meter traceable to NIST primary flow standards. Section 3 describes the RATA method; Section 4 gives a physical explanation why the X-pattern CEMS compensates for swirl and a single path USM does not; Section 5 presents the results of an S-probe RATA, a single path USM, and the X-pattern CEMS tested in the SMSS facility; and Section 6 states our conclusions.

2 Scale-Model Smokestack Simulator (SMSS)

NIST designed and built the Scale-Model Smokestack Simulator (SMSS) shown in Figs. 2 and 3. The SMSS is used to 1) assess the flow measurement accuracy of the Relative Accuracy Test Audit (RATA), 2) quantify the performance of Continuous Emissions Measurement Systems (CEMS) flow monitors, and 3) serve as a testbed for new stack flow measurement techniques. In this work, we used the SMSS to evaluate the performance of the S-probe RATA, single-path CEMS flow monitor, and X-pattern CEMS flow monitor. Because stack flow measurements depend primarily on the characteristics of the velocity field, and not on the gas composition, the SMSS uses air as a surrogate for flue gas. In its Test Section (D_{test} = 1.2 m diameter), the SMSS generates complex velocity fields (*i.e.*, turbulent, swirling flow, with an asymmetric velocity profile) prevalent in industrial-scale stacks.

Figure 2. Scale-Model Smokestack Simulator (SMSS) showing the sections where air enters and exits the facility. RATA and CEMS performance evaluation are performed inside the building in its 1.2 m diameter Test Section.

Air enters the SMSS at the Air Intake Unit shown in Fig. 2. In the Test Section, the average air flow velocity ranges from 5 m/s to 26 m/s; this range spans the velocities in industrial scale CFPPs stacks. The unique feature of the SMSS is that the average flow velocity (V_{NIST}), is traceable to NIST's primary flow standards and known to 0.7 % expanded uncertainty [14, 15]. Therefore, the SMSS can assess the *absolute accuracy* of CEMS flow monitors. In contrast, typical RATA measurements can assess only the *relative accuracy* of CEMS systems because they rely on S-probes that are often not calibrated and that do not account for the complexity of the flow field.

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

The schematic diagram shown in Fig. 3 highlights the functional design of the SMSS facility. Quiescent air is drawn into the air intake unit by fans at the facility exit. Cross-flow velocity components are damped as air moves through the intake unit. Low-speed air exiting the intake unit accelerates through a cone (Fig. 1) and establishes a low-swirl, nearly-uniform velocity profile in the Reference Section. We measure the flow in the Reference Section using a NIST-calibrated, 8 path ultrasonic flow meter (USM) [15]. Because the velocity field in the Reference Section is practically free of flow distortions, a high accuracy flow measurement is straightforward. We point out, however, that the 8 path USM can accurately measure the flow even when the flow has significant levels of swirl and asymmetry [16, 17].

In contrast, to the distortion-free flow in the Reference Section, the velocity field in the Test Section is designed to replicate the complex velocity fields typical to stack flows. Flow distortions in an industrial scale smokestack are caused by fans, reducers, and most notably a sharp corner where the flue gas enters the stack. In a similar manner, flow distortions in the SMSS Test Section result from the sharp corner shown in Fig. 3.

The leaks in the SMSS are negligible; therefore, the air flux in the Test and Reference Sections are equal. The NIST-traceable flow measurement in the Reference Section determines the average flow velocity (V_{NIST}) in the Test Section by accounting for the diameter and gas density changes between the sections. V_{NIST} is used to assess the performance of CEMS flow monitors and the flow RATA.

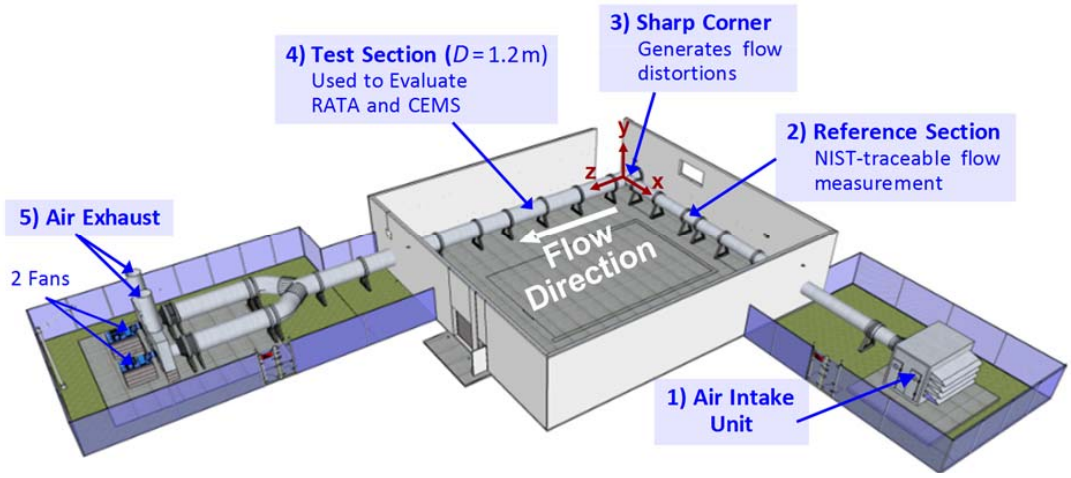


Figure 3. Schematic of the 5 sections of the SMSS facility. 1) air enters the intake unit, 2) a well-conditioned flow is measured in the Reference Section using a NIST-traceable flow meter, 3) sharp corner generates swirling, asymmetric flow, 4)Test Section used for assessing the accuracy of RATA and CEMS in distorted flows, and 5) fans that drive the flow and exhaust it to the atmosphere.

3 S-probe RATA Methodology

We conducted an S-probe RATA 12 diameters (14.4 m) downstream of the sharp corner shown in Fig. 3. During setup, an S-probe was installed in the SMSS Test Section with its positive pressure port aligned with the pipe centerline (*i.e.*, the z-axis in Fig. 3). The axial velocities were measured at 2*N* points along the two orthogonal chords shown in Fig. 1B. In accordance with EPA

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Method 2 [9], the N points on each chord were spaced at the centroids of equal area so that the flow velocity measured by the S-probe is

$$V_{\text{Sprobe}} = \sum_{n=1}^{2N} w_n V_{\text{Sprobe,n}} \quad , \tag{2}$$

where the weighting factor is $w_n = 1/2N$, and the S-probe velocity measured at each n^{th} point is deduced from the Bernoulli equation [18]

$$V_{\text{Sprobe,n}} = C_p \sqrt{2\Delta P_n / \rho}\cos(\beta_{\text{FLOW,n}}) \quad , \tag{3}$$

where C_p = 0.84 is the value of the S-probe calibration factor conventionally used in the stack testing community. In Eq. (3), ρ is the average air density in the RATA cross section; it is determined to better than 0.2 % using an equation of state [19] that depends on the measured pressure, temperature, and relative humidity. The flow angle is $\beta_{\text{FLOW,n}}$ and ΔP_n is the measured differential pressure with the S-probe oriented at the flow angle. As illustrated in Fig. 4, the flow angle is determined by first rotating the probe to the yaw-null angle ($\beta_{\text{NULL,n}}$) where ΔP_n = 0, and then rotating the S-probe 90 ° in the negative yaw angle direction.

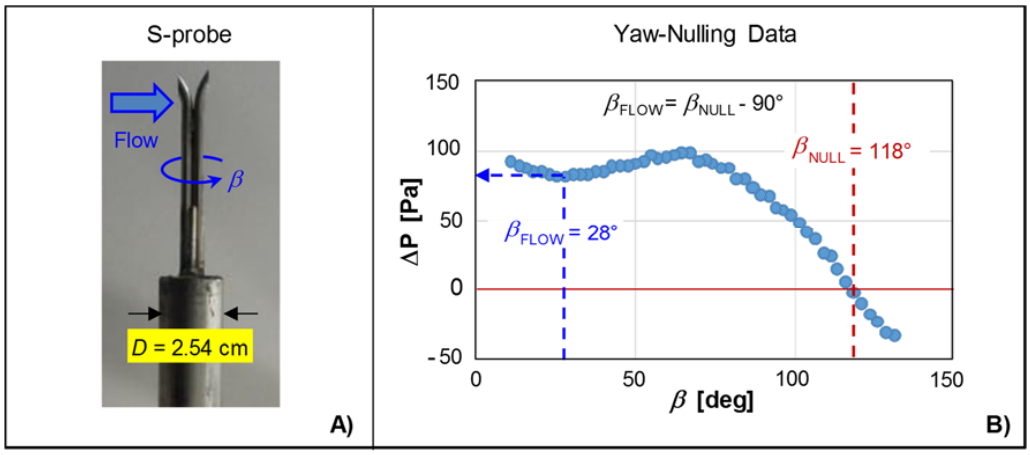


Figure 4. S-probe yaw-nulling method: A) S-probe showing the direction positive yaw angle rotation (β), and B) measurements of the differential pressures (ΔP) across S-probe ports as a function of β. The null angle (βNULL) is characterized by ΔP = 0 and βFLOW is 90 ° from βNULL.

4 CEMS Measurements

We made flow measurements using 3 single-path USMs installed 10 diameters downstream from the sharp corner in Fig. 3. All 3 USMs were installed in the same flow meter body shown in Fig. 5A. Figures 5B, 5C, and 5D show the orientation of each path in the meter body. Path 1 is vertically oriented at a 45 ° angle relative to the Test Section's centerline while Paths 2 and 3 are oriented horizontally at a 45 ° angle relative to the centerline. The linear distance between the transducers is the path length (L_p), and the orientation of the path relative to the pipe centerline is the path angle (φ). In industrial CFFP stacks, L_p and φ are not always accurately measured. Instead, the flow RATA

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

calibration corrects 1) for biases attributed to CEMS dimensional errors and 2) for flow related biases attributed to the complex stack velocity field. Since the goal of this work is to assess the absolute accuracy of a single-path and the X-pattern USM, we used a laser tracking system to measure L_p's and φ's with expanded uncertainties of 0.25 % and 0.5 °. The uncertainties of these dimensional measurements are negligible relative to flow related uncertainties attributed to profile and swirl effects; therefore, differences between the average velocity measured by the single path and X-pattern relative to the NIST standard velocity can be ascribed to flow related errors.

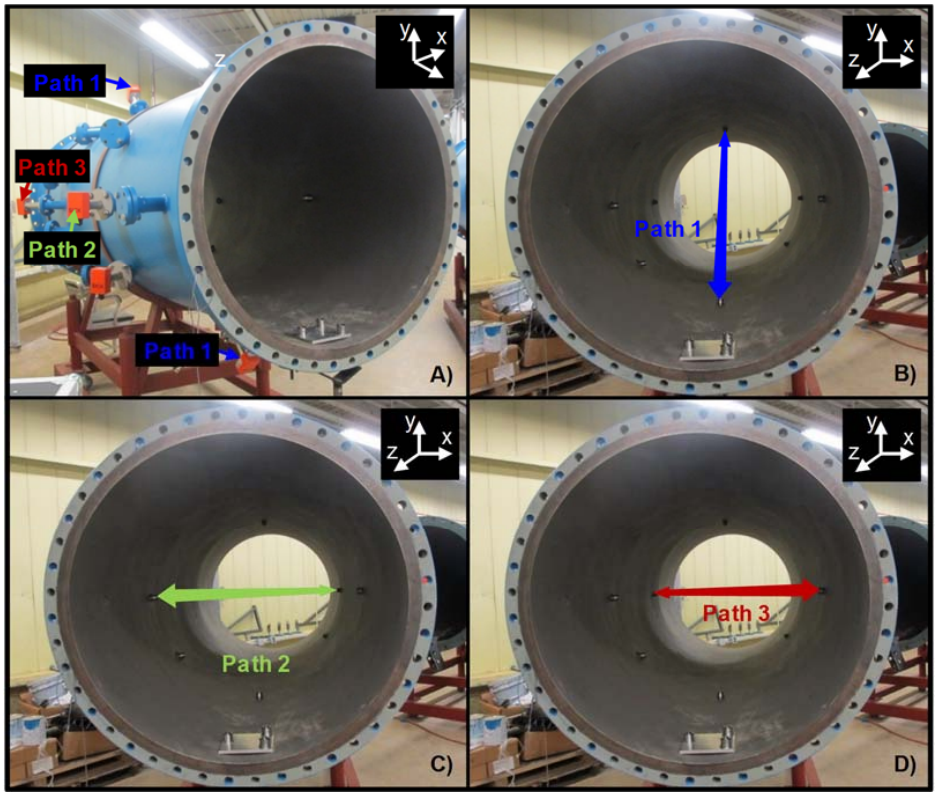


Figure 5. Three single-path USMs installed in the same spool located 10 D_{test} downstream of sharp corner in Fig. 3: A) Three transmitting/receiving transducers of the USMs on the pipe exterior. B), C), and D) are interior views of the 3 paths used by the 3 transducer pairs. Paths 2 and 3 together comprise an X-pattern USM.

4.1 USM Principle of Operation

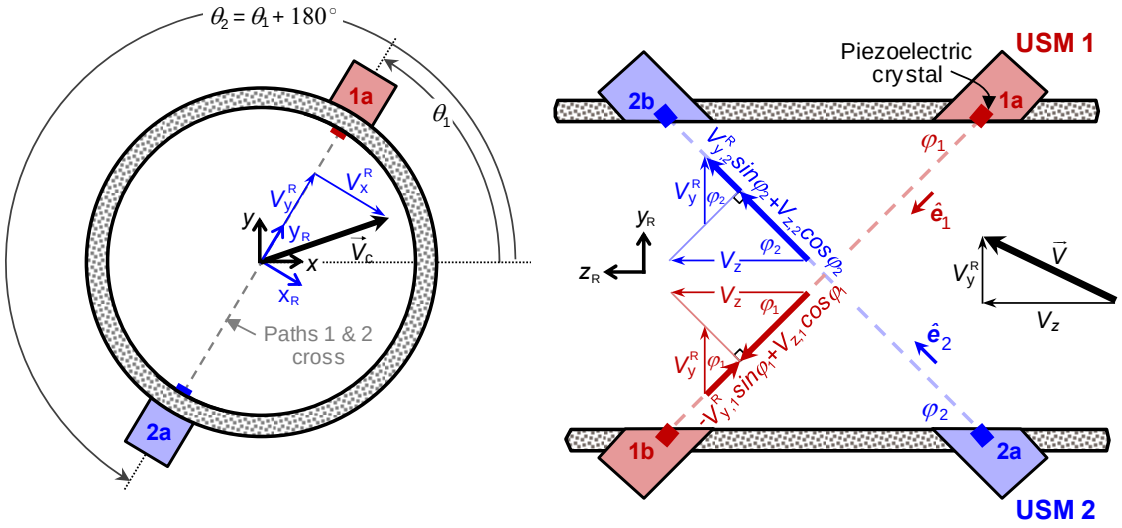
Industrial USMs used in stack applications measure the time it takes an ultrasonic pulse to travel the distance L_p. Two distinct time measurements are necessary to determine the flow velocity. The 2 times are denoted t_w and t_a, respectively, where t_w is the time it takes an ultrasonic pulse to travel L_p *with* (or assisted by the flue gas velocity) and t_a is the time required for another pulse to travel L_p moving *against* (or retarded by the velocity of the flue gas). If a_p and $\overline{V}_{USM,p}$ are the respective speed of sound along the acoustic path and path velocity, then the two measured times are 1) $t_w = L_p/(a_p + \overline{V}_{USM,p})$ and 2) $t_a = L_p/(a_p - \overline{V}_{USM,p})$. When the flow along the path moves in the same

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

direction as the ultrasonic pulse, the pulse moves at the sound speed augmented by $\bar{V}_{USM,p}$. Likewise, when the fluid moves opposite to the pulse, the pulse speed is the sound velocity reduced by $\bar{V}_{USM,p}$. Solving these two equations for the path velocity gives

$$\bar{V}_{USM,p} = \frac{L_p}{2}\left(\frac{1}{t_w} - \frac{1}{t_a}\right). \quad (4)$$

For low speed flows ($M_p < 0.1$) where $M_p = V_p/a_p$ is the Mach number, Eq. (4) is accurate to the order M_p^2. Given that most stack flows satisfy this Mach number criterion, USMs can typically measure $\bar{V}_{USM,p}$ to better than 0.5 %, provided L_p is known, and the USM signal processing is sufficiently robust to accurately measure t_a and t_w. However, the bulk flow is determined by the component of velocity along the stack axis, not by $\bar{V}_{USM,p}$. Common practice is to estimate the axial velocity by dividing $\bar{V}_{USM,p}$ by cos(φ) as shown in Eq. (1). This approximate formulation accurately predicts the flow velocity only when the magnitude of the cross-flow velocity (V_c) is small relative to the axial flow and profile effects are negligible. If V_c is significant relative to the flow velocity, $\bar{V}_{USM,p}$ will consist of velocity components from flow along the stack axis (V_z), but also from the cross flow (V_c). That is, V_c will influence the transit time of an ultrasonic pulse moving from the emitting to the receiving transducers. Since stack flows usually have significant swirl levels, single-path USMs generally do not provide accurate flow measurements. In contrast, the X-pattern design compensates for swirl-related errors.



A) Cross Sectional View B) Symmetry Plane in Rotated Coordinate System

Figure 6. Sketch showing 2 single-path USMs 1 and 2, which together comprise the X-pattern configuration. A) cross-sectional view displaying USM installation angles θ_1 and θ_2, and cross-flow velocity $\vec{V}_c$ in 2 coordinate frames (x, y) and (x_R, y_R), and B) shows a symmetry plane in rotated coordinates where the velocities along paths 1 and 2 have contributions both from the axial and cross-flow components of the fluid velocity ($\vec{V}$).

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

4.2 X-pattern Swirl Compensation

An X-pattern consists of the two USMs 1 and 2 installed in the same plane with opposite orientations (in Fig. 6A, $\theta_2 = \theta_1 + 180°$.) and with the same path angle, $\varphi_1 = \varphi_2$. Here, we show that perfect compensation for swirl (*i.e.*, the non-axial velocity components) occurs only where the two ultrasonic beams cross on the axis of the stack at $y_R = 0$. At other y_R locations the swirl cancelation is not perfect; however, extensive empirical evidence has shown the X-pattern does an excellent job compensating for swirl [20]. We give an intuitive picture and a physical explanation on why the X-pattern USM provides superior swirl compensation relative to the currently used single path USM.

Figure 6 shows an orthographic drawing of an X-pattern USM. In this example, the fluid velocity vector is $\vec{V} = V_z\vec{k} + \vec{V}_c$ where V_z is the axial velocity, and $\vec{V}_c$ is the cross-flow velocity vector that generates swirl errors. Figure 6A shows the components of $\vec{V}_c$ in the coordinate frame (x_R, y_R, z_R) rotated about the z-axis so that V_x^R is orthogonal to acoustic paths 1 and 2, and does not influence the USM transit time measurements. In the rotated reference frame, the acoustic path trajectories are described solely by the coordinates y_R and z_R as illustrated in Fig. 6B. Therefore, the travel times of an ultrasonic pulse moving along paths 1 or 2 are dependent only on the cross-flow velocity V_y^R and on the axial velocity V_z. Summing the contributions from both velocity components gives the following average velocity along path 1

$$\overline{V}_{USM,p1} = \overline{V}_{z,1}\cos\varphi_1 - \overline{V}^R_{y,1}\sin\varphi_1 \tag{5a}$$

where $\overline{V}_{z,1}$ and $\overline{V}^R_{y,1}$ are the average axial and cross-flow velocities along path 1. Likewise, the average velocity along path 2 is

$$\overline{V}_{USM,p2} = \overline{V}_{z,2}\cos\varphi_2 + \overline{V}^R_{y,2}\sin\varphi_2 \tag{5b}$$

where $\overline{V}_{z,2}$ is the average axial flow velocity along the path and $\overline{V}^R_{y,2}$ is the average cross-flow velocity along path 2[4] Due to the orientation of paths 1 and 2 a positive cross-flow velocity reduces the speed on path 1, but increases the speed on path 2. Using Eq. (1) to calculate the flow velocity on these paths leads to

$$\overline{V}_{USM,f1} = \frac{\overline{V}_{USM,p1}}{\cos\varphi_1} = \overline{V}_{z,1} - \overline{V}^R_{y,1}\tan\varphi_1 \tag{6a}$$

for path 1 and

$$\overline{V}_{USM,f2} = \frac{\overline{V}_{USM,p2}}{\cos\varphi_2} = \overline{V}_{z,2} + \overline{V}^R_{y,2}\tan\varphi_2\,, \tag{6b}$$

4 Equations (6a) and (6b) can also be developed taking the dot product of the fluid velocity vector with the unit tangent vectors $\hat{e}_1$ and $\hat{e}_2$, respectively: $\overline{V}_{USM,p1} = \vec{V}\cdot\hat{e}_1$ and $\overline{V}_{USM,p2} = \vec{V}\cdot\hat{e}_2$ where based on the geometry in Fig. 6B $\hat{e}_1 = -\sin\varphi_1\hat{e}^R_y + \cos\varphi_1\hat{k}$ and $\hat{e}_2 = \sin\varphi_2\hat{e}^R_y + \cos\varphi_2\hat{k}$ and the unit vector in the y_R direction is $\hat{e}^R_y = \cos\theta_1\hat{i} + \sin\theta_1\hat{j}$.

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

for path 2. These equations illustrate how the cross-flow velocities $\overline{V}^{R}_{y,1}$ and $\overline{V}^{R}_{y,2}$ cause errors in the performance of single-path USMs. The flow velocity of USM 1 is decreased by $\overline{V}^{R}_{y,1} tan\varphi_1$ while it is increased by $\overline{V}^{R}_{y,2} tan\varphi_2$ for USM 2. Since the X-pattern is the arithmetic average of USM 1 and 2 its flow velocity is given by

$$\overline{V}_{USMX,f} \approx \frac{\overline{V}_{z,1} + \overline{V}_{z,2}}{2} \tag{6c}$$

where the term $(\overline{V}^{R}_{y,2} tan\varphi_2 - \overline{V}^{R}_{y,1} tan\varphi_1)/2$ has been omitted from Eq. (6c) since its value is nearly zero. The X-pattern path angles are the same ($\varphi_2 = \varphi_1$), and the average velocities along each path tend to be of the same sign and magnitude ($\overline{V}^{R}_{y,2} \approx \overline{V}^{R}_{y,1}$).

Swirl is dissipated by viscous shear caused by the no-slip condition at the pipe wall. For high Reynolds number (*Re*) flows like the SMSS (*Re* = 4.7×10^5 to 2.0×10^6) and CFPPs stacks (*Re* = 2.4×10^6 to 2.6×10^7), experiments show that swirl effects can persist for 100 pipe diameters or more [21, 22]. In contrast, the axial distance between paths 1 and 2 of the X-pattern in Fig. 6B is less than one pipe diameter ($z_{R2} - z_{R1} < D_{test}$) at any fixed value of y_R. The local cross-flow velocity at a fixed value of y_R is $V^{R}_{y,1}(x_R, y_R, z_{R1})$ at path 1 and $V^{R}_{y,2}(x_R, y_R, z_{R2})$ at path 2. Since swirl requires many pipe diameters to dissipate, the local cross-flow velocities are nearly equal on acoustic paths 1 and 2, $V^{R}_{y,1}(x_R, y_R, z_{R1}) \approx V^{R}_{y,2}(x_R, y_R, z_{R2})$. Consequently, the average cross-flow velocities are also nearly equal, $\overline{V}^{R}_{y,2} \approx \overline{V}^{R}_{y,1}$. Therefore, the X-pattern compensates for swirl errors and generally outperforms the single-path USM. The effectiveness of the X-pattern to dissipate swirl has been empirically and computationally verified by numerous researchers [20 ,23 -25].

5 Results from the SMSS

5.1 Velocity Profile and Cross Flow in SMSS Test Section

We used an S-probe to measure the normalized axial velocity profile ($V_{Sprobe,n}/V_{NIST}$) and the flow angle (β_{FLOW}) along two orthogonal chords in the SMSS Test Section. Measurements were made at z = 12 D_{test} where z the axis centerline originating at the coordinate system located at the sharp corner shown in Fig. 3. We measured the axial velocity and flow angle along the chords oriented at θ_1 = 45 ° and at θ_2 = 135 ° using a custom designed automated traversing system (ATS) [14]. Figure 1B shows the orientation of the two traverse chords relative to the xy-coordinate axes. Although the figure shows two S-probes simultaneously traversing the cross section, the ATS (not shown) only has one access port. Once the S-probe is installed into this access port the ATS rotates the pipe section containing the S-probe about the z-axis to the specified θ orientations. At each θ orientation, the traverse system moves the probe across the chord to the prescribed traverse points. At each point, the ATS rotates the S-probe about its axis to find β_{NULL} and β_{FLOW}. The traverse distance at any point on the chord is defined by the parameter ξ, which is zero at the inner pipe wall (*i.e.*, the wall closest to the ATS assess port) and *D* at the outer pipe wall (farthest from the ATS

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

assess port). We measured $V_{Sprobe,n}/V_{NIST}$ and β_{FLOW} at 24 points located at the centroids of equal area on each chord. At each point the S-probe velocity is determined using Eq. (3).

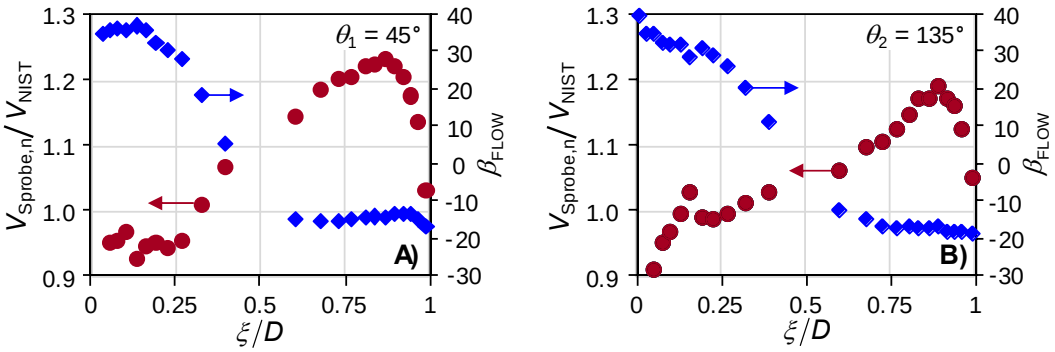


Figure 7. Normailized velocity profile (●) and flow angle (◆) as functions of the dimensionless distance ξ/D along 2 orthognal chords located 12 diameters (D = 1.2 m) downstream from the sharp corner in SMSS. A) 24-point S-probe traverse at θ_1 = 45 °, and B) 24-point S-probe traverse θ_2 = 135 °.

The results of the S-probe traverse are shown in Fig. 7 where the circles (●) are the normalized velocity profiles and the diamonds (◆) are the flow angle. During the traverse the flow velocity was maintained at V_{NIST} = 5.27 m/s. Figure 7A shows the results for θ_1 = 45 ° and Fig. 7B shows the results for θ_2 = 135 °. The axial velocity profiles are both skewed toward the outer pipe wall; however, the skew exhibited on the velocity profile at θ_1 = 45 ° is slightly more pronounced. The sharp decrease in velocity near the inner and outer walls indicate the thickness of the boundary layer. The cross-flow velocity in the SMSS Test Section is substantial as indicated by the flow angle ranging from near 40 ° at the inner wall to almost −20 ° at the outer wall in both orientations.

5.2 S-probe RATA in the SMSS

With the traverse system installed at z = 12 D we assessed the accuracy of an S-probe RATA on 2 diametric chords oriented at θ_1 = 45 ° and θ_2 = 135 °. We followed the common RATA protocol that uses the value C_p = 0.84 for the S-probe calibration factor. The RATA was done at two flow velocities: 5.28 m/s and 23.29 m/s. These velocities approximate low and high loads in a CFPP. At each flow, we conducted a 12-point RATA, a 24-point RATA, and a 48-point RATA. On each chord, the traversing system moved the S-probe radially to the specified RATA point and performed the yaw-nulling procedure described in Section 3 to find $\beta_{FLOW,n}$. The axial velocity at each point was calculated using Eq. (3) and the flow velocity was calculated using Eq. (2). The results are shown in Table 1.

At both the high and low loads, the S-probe RATA overpredicted the actual flow in the SMSS. The largest error of 6.1 % occurred at the low load for the 12-point RATA. The modest improvements for the 24-point and 48-point RATA probably resulted from improved resolution of the sharp velocity gradient in the boundary layer close to the wall. The conventional value (C_p = 0.84) of the S-probe calibration factor may be too large; if so, it is partly responsible for overpredicting the actual flow velocity. We suspect that significant levels of pitch, which could not be measured with the S-probe, also contribute to the overprediction. Independent of the causes, the overprediction suggests that the quantity of hazardous emissions from actual CFPPs might be significantly overestimated; however, this suggestion must be tested in actual

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

CFPP stacks by comparing conventional RATA measurements with rigorous measurements traceable to flow standards.

Table 1. ***S-probe flow RATA performed 12 D downstream of the sharp corner in SMSS facility. Measurements taken by traversing S-probe along orthogonal chords oriented at θ_1 = 45 ° and θ_2 = 135 ° at low and high loads with C_p = 0.84.***

No. of Points [ ]	V_{NIST} [m/s]	V_{Sprobe} [m/s]	$100\left[\frac{V_{Sprobe}}{V_{NIST}} - 1\right]$ [%]
12	5.28	5.6	+6.1
24	5.28	5.57	+5.5
48	5.28	5.53	+4.7
12	23.29	24.83	+6.6
24	23.29	24.68	+6.0
48	23.29	24.47	+5.1

5.3 Single-Path and X-Pattern CEMS in the SMSS

We assessed the accuracy of the 3 single-path USMs shown in Fig. 5 and an X-Pattern design comprising paths 2 and 3 together. Unlike industrial CEMS applications, we did not calibrate the USMs *via* the RATA. Instead we computed the flow velocities using only the raw meter transit times t_a and t_w along with the dimensionally measured path length (L_p) and path angle (φ). First we calculated the path velocity ($\overline{V}_{USM,p}$) using Eq. (4), and then divided it by cos(φ), as specified by Eq. (1) to determine the flow velocity ($\overline{V}_{USM,f}$). For the X-pattern we computed the flow velocity by averaging the flow velocities from USM 2 and 3, $\overline{V}_{USMX,f} = (\overline{V}_{USM2,f} + \overline{V}_{USM3,f})/2$. The accuracy was assessed by comparing the respective USM flow velocities to the NIST traceable average velocity (V_{NIST}).

The results for the 3 single-path USMs and the X-pattern are shown in Fig. 8. The normalized flow velocity ($\overline{V}_{USM,f}/V_{NIST}$) is plotted against V_{NIST}. Ideally, the normalized flow velocity would be unity, indicating that the USM is not affected by swirl or profile distortions. All three single-path USMs showed substantial biases. USM 1 denoted by the triangles (▲) was biased by nearly 6 % over the entire flow range. USM 2 indicated by the diamonds (◆) was also biased high, but it had a larger error ranging from 14 % to 17 % depending on flow. In contrast, USM 3 denoted by the squares (■) was biased low by 14 % to 17 %. Based on the approximate swirl compensation of the X-pattern design (see Section 4.2) we were not surprised that the biases of USM 2 and 3 had opposite signs. However, we were surprised at the remarkable agreement between the X-pattern denoted by the circles (●) and V_{NIST}. Over the entire range of flow the agreement was better than 0.5 %. This sub-one-percent agreement was unexpected because the X-pattern compensates for swirl, but it does not compensate for profile errors. More testing is required to assess the sensitivity of this result to the installation angle (θ), and to different flow installations. Nevertheless, these initial results indicate that the X-pattern USM is significantly more accurate than a single-path USM.

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

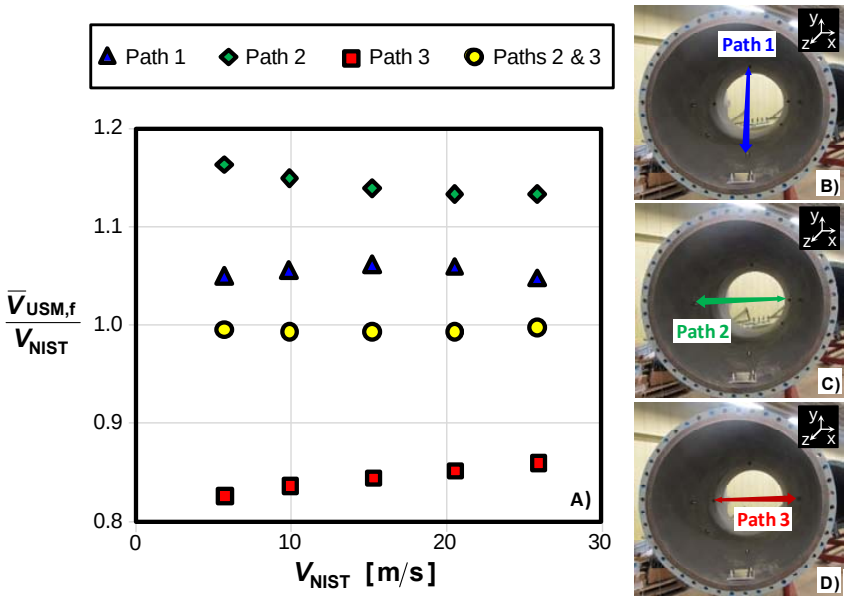


Figure 8. Data comparing the accuracy of 3 single-path USMs versus an X-pattern design. A) plot of USM flow velocity normalized by V_{NIST}, B) orientation of path 1, C) orientation of path 2, and D) orientation of path 3. (Note that paths 2 and 3 together comprise the X-pattern USM).

6 Conclusions

We used the SMSS to generate a known flow (V_{NIST}) with an uncertainty of 0.7 % (at 95 % confidence level) with swirl and profile asymmetry in its D =1.2 m Test Section. The axial velocity and yaw angle along two diametric chords showed that the velocity was ~30 % larger on one side of the Test Section than the other. Cross-flow was substantial; the yaw angle varied from ~40 ° on one side of a chord to −20 ° on the other side. This complex velocity field simulates high-distortion flow condition in a CFPP stack.

In SMSS Test Section, we determined the flow using an S-probe RATA, three single-path USMs oriented at a different installation angles, and an X-pattern USM. A 12-point S-probe RATA overpredicted the actual flow velocity (V_{NIST}) by as much as 6.1 %. The level of overprediction decreased to 4.7 % upon increasing the number of points to 48. The remaining overprediction is probably a consequence of the S-probe's inability to compensate for non-zero pitch and the default calibration factor of 0.84. To further improve the RATA's accuracy, we recommend replacing S-probes with 3 D probes that measure the entire velocity vector.

The 3 single-path USMs had flow errors ranging from 5 % to 17 % of V_{NIST} depending on the installation angle. In contrast, the X-pattern USM determined flows within 0.5 % of V_{NIST} over the entire flow range. We attribute this remarkable result to the swirl compensation of the X-pattern design. Because the X-pattern does not compensate for velocity profile effects, additional testing

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

is necessary to understand the sensitivity of the X-pattern to asymmetry. We plan to perform these tests both in the SMSS facility and in a full-scale CFFP stack.

These initial results show that the errors in an S-probe flow RATA can (at least for certain flows) exceed the errors of an X-pattern USM monitor by nearly an order of magnitude. If future research confirms this observation, improved RATAs will require replacing S-probes with more advanced pitot probes that complement the accuracy offered by the X-pattern.

REFERENCES

[1] Dorko, W. D., Kelley, M. E., and Guenther, F. R., ***The NIST Traceable Reference Material Program for Gas Standards***, NIST Special Publication 260-126 Rev 2013, 2015.

[2] U.S. Environmental Protection Agency. ***U. S. EPA Ambient Air Monitoring Protocol Gas Verification Program Annual Report CY 2010***, EPA Publication No. EPA-454/R-11-005, 2011.

[3] Hatfield, J., Pierce, G., and Beusse, R., ***EPA Needs an Oversight Program for Protocol Gases Report No. 09-P-0235***, Sept. 16, 2009.

[4] Gameson, L., Carney, J., Hodges, J. T., and Guenther, F. R., ***Environmental Protection Agency Blind Audit 2013***, Report of Analysis No. 646.03-14-071b, 2015.

[5] Gameson, L., and Guenther, F. R., ***Environmental Protection Agency Blind Audit 2010***, Report of Analysis No. 639.03-11-026a, 2011. NIST Report of Analysis No. 639.03-11-026a, 2011.

[6] Norfleet, S. K., and Roberson, W. R., ***Part 75 CEMS Equipment - What's Everyone Using? A Look at Current Trends in CEMS Providers***, EPRI CEMS Users Group Meeting, San Diego, CA, May 2003.

[7] Environmental Protection Agency***, 40 CFR Part 60 Application Method 1 Traverse Points – Stationary Sources***, Washington D.C., 1996.

[8] Environmental Protection Agency***, 40 CFR Part 60 Application Method 2 Velocity and S-type Pitot***, Washington D.C., 1996.

[9] Environmental Protection Agency, ***40 CFR Part 60 Application Method 2G Stack Gas-Velocity/Volumetric Flow Rate-2D probe***, Washington D.C., 1996.

[10] Shinder, I. I., Khromchenko, V. B., and Moldover, M. R., ***NIST's New 3D Airspeed Calibration Rig, Addresses Turbulent Flow Measurement Challenges,*** 9th International Symposium on Fluid Flow Measurement (ISFFM), Washington, DC, U. S., April 2015.

[11] Norfleet, S. K., Muzio L. J., and Martz T. D., ***An Examination of Bias in Method 2 Measurements Under Controlled Non-Axial Flow Conditions***, Report by RMB Consulting and Research, Inc., May 22, 1998.

[12] Borthwick, R. P., Whetstone, J., Yang, J., and Possolo, A., ***Examination of United States Carbon Dioxide Emission Databases***, EPRI CEM User Group, Chicago, IL: June 2011.

[13] Quick, J. C., ***Carbon dioxide emission tallies for 210 U.S. coal-fired power plants: A comparison of two accounting methods,*** Journal of the Air & Waste Management Association, January 2014.

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

[14] Johnson, A. N., Boyd, J., Harman, E., Khalil, M., Ricker, J. E., Crowley, C. J., Bryant, R. A.; and Shinder, I. I., ***Design and Capabilities of NISTs Scale-Model Smokestack Simulator (SMSS)***, 9th International Symposium on Fluid Flow Measurement (ISFFM), Arlington, VA, US, April 2015.

[15] Johnson, A. N., Harman, E., and Boyd, J. T., ***Blow-Down Calibration of a Large 8 Path Ultrasonic Flow Meter under Quasi-Steady Flow Conditions***, The 16th International Flow Measurement Conference FLOMEKO 2013, Paris France, September 2013.

[16] Brown, G. J., Freund, W. R., and McLachlan, A., ***Testing of an 8-Path Ultrasonic Meter to International Standards with and without Flow Conditioning***, AGA Operations Conference, Orlando, FL, May 21 - 24, 2013.

[17] Brown, G. J., Augenstein D. R., Cousins, T., ***8-Path Ultrasonic Master Meter for Oil Custody Transfer***, XVIII IMEKO World Congress Metrology for a Sustainable Development, Rio de Janeiro, Brazil, September, 17 - 22, 2006.

[18] Cimbala, J., and Cengel, Y. A., ***Fluid Mechanics: Fundamentals and Applications***, McGraw-Hill series in mechanical engineering, 2006.

[19] Jaeger, K. B. and Davis, R. S., ***A Primer for Mass Metrology, National Bureau of Standards (U.S.), Special Publication 700-1***, Industrial Measurement Series. 1987: 79.

[20] Drenthen, J. G., Kurth M., Hollander H., Vermeulen, M., ***Reducing installation effects on ultrasonic flow meters***, 7th International Symposium on Fluid Flow Measurement (ISFFM), Anchorage, AK, August 12-14, 2009.

[21] Mattingly, G. E. and Yeh, T. T., ***Effects of pipe elbows and tube bundles on selected types of flowmeters***, Flow Measurement and Instrumentation, Vol. 2, Number 1, p. 4, 1991.

[22] Mattingly, G. E. and Yeh, T. T., ***Summary Report of NIST's Industry-Government Consortium Research Program on Flowmeter Installation Effects with Emphasis on the Research Period November 1988 to May 1989***, U.S. Department of Commerce, NISTIR 4310, 1989.

[23] Brown, G. J., and Freund, W. R., and McLachlan, A., ***Analysis of Diagnostic Data from an 8-Path Ultrasonic Meter Testing of an 8-Path Ultrasonic Meter***, AGA Operations Conference, Orlando, FL, May 21 – 24, 2013.

[24] Zhang, L., Heming, H., Tao, M., and Chi., W., ***Effect of Flow Disturbance on Multi-path Ultrasonic Flowmeters*****,** FLOMEKO, Paris, France, September 24-26, 2013.

[25] Brown, G., and Coull, C., and Barton, N., ***Installation Effect on Ultrasonic Flowmeters and Evaluation of Computation Fluid Dynamics Prediction Methods*****,** 4th International South East Asia Hydrocarbon Flow Measurement Workshop, Kuala Lumpur, Malaysia, March 9–11 2005.

Boyd, Joey; Filla, Bernard; Johnson, Aaron; Moldover, Michael; Shinder, Iosif.
"Progress Towards Accurate Monitoring of Flue Gas Emissions."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Nanoscale 3D Shape Process Monitoring Using TSOM

Ravi Kiran Attota
Engineering Physics Davison, PML, NIST, Gaithersburg, MD 20899, USA
Author e-mail address: Ravikiran.attota@nist.gov

Abstract: Through-focus scanning optical microscopy (TSOM) is sensitive to three-dimensional shape changes of nanoscale to microscale targets. Here we demonstrate process monitoring method of 3D targets using TSOM down to sub-nanometer.
OCIS codes: (120.3930) Metrological instrumentation; (170.0110) Imaging systems; (110.0180) Microscopy

1. Introduction

Through-focus scanning optical microscopy (TSOM) is a novel method [1-5] that allows conventional optical microscopes to collect dimensional information down to the nanometer level by combining 2D optical images captured at several through-focus positions, transforming conventional optical microscopes into truly 3D metrology tools for nanoscale to microscale dimensional analysis with nanometer scale sensitivity. TSOM identifies the 3D profile dimensional changes at the nanoscale for small perturbations in sidewall angle, width and height of isolated lines with little or no ambiguity. This work will present TSOM results to demonstrate the metrology application to features below 50 nm, showing the ability to measure changes in line width (LW) line height (LH), and sidewall angle (SWA) variations.

2. Results

Here we present experimental TSOM results using isolated Si lines on a Si substrate. The designed line widths varying between 30 nm and 40 nm were fabricated using standard fabrication methods. The line widths were carefully measured using a critical dimension atomic force microscope (CD-AFM). The smallest and the largest measured linewidths were 44.2 nm and 55.3 nm, respectively, with a fairly uniform height of 71.4 nm. Standard deviation values for line width and height are 1.6 nm, and 0.4 nm, respectively. A typical TSOM image of the 44.2 nm wide line is shown Fig. 1(a) (using 546 nm illumination wavelength, 0.8 numerical aperture (NA) and 0.27 illumination NA). Intensity profiles at several focus positions shown on the right-side help visualize the color scheme in the TSOM image and also demonstrate the way profile changes occur with focus position.

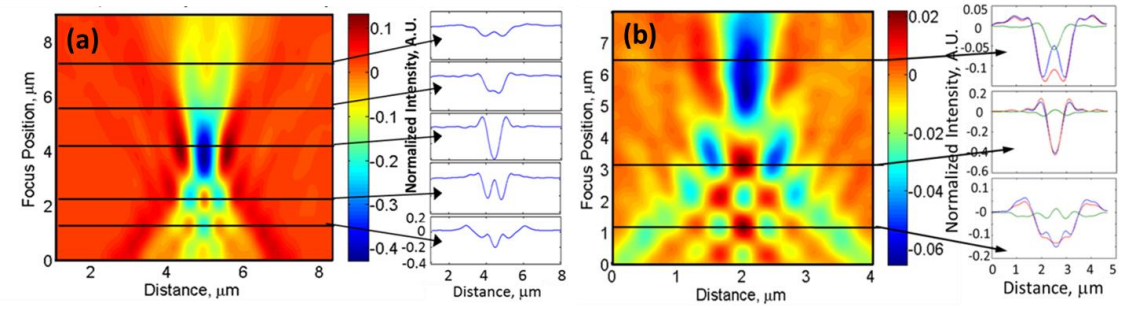


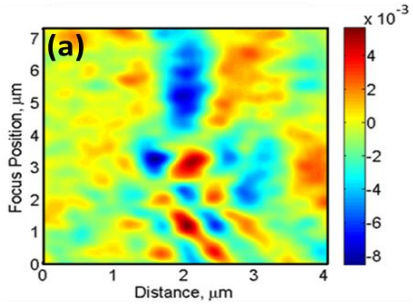
Fig. 1 Intensity of (a) normalized TSOM image of a 44.2 nm wide line (left) with intensity profiles at the selected focus positions on the right side, (b) a normalized differential TSOM image obtained using 44.2 nm and 55.3 nm wide lines (left) with intensity profiles at the selected focus positions and the corresponding differential profiles on the right side. The blue profile is for the 44.2 nm line, the red profile is for the 55.3 nm line, and the green profile is the differential.

A differential TSOM (D-TSOM) image is a pixel-by-pixel difference between two TSOM images. A D-TSOM image highlights nanoscale 3D shape differences. A D-TSOM image for 11.1 nm difference in the line width is shown in Fig. 1(b), which has an optical intensity range (OIR) of 8.5. OIR indicates the optical signal strength. On

Attota, Ravikiran.
"Nanoscale 3D Shape Process Monitoring Using TSOM."
Paper presented at Applied Industrial Optics, Orlando, FL, United States. June 25, 2018 - June 28, 2018.

the right side, the optical intensity profiles and the differential profile from the two TSOM images are shown for the three selected focus positions. The maximum difference occurs in the upper region of the D-TSOM image, away from the best focus position (highest image contrast position) which is at about a 4 μm focus position.

The TSOM method was able to detect a linewidth difference as small as 0.8 nm as shown in Fig. 2(a), which has an OIR of 1.5. This is still above the noise threshold of about one, demonstrating the experimental feasibility of detecting sub-nanometer differences in linewidths using the TSOM method.



(b)

Nom.	λ	Minimum Sensitivity		
LW		LW	SWA	LH
nm	nm	nm	Degree	nm
22	546	0.04	0.06	0.08
22	248	0.21	0.38	0.15
22	193	0.08	0.1	0.04
16	546	0.07	0.26	0.24
16	258	0.12	0.45	0.2
16	193	0.13	0.27	0.15

Fig. 2 (a) Intensity of a D-TSOM image for 0.8 nm difference in the LW (47.9 nm and 48.7 nm). (b) Minimum detectable dimensional differences using the TSOM method for the three illumination wavelengths (λ).

Optical simulations were used to predict how TSOM would perform for measuring structural perturbations for nominally 22 nm and 16 nm wide line. The simulations used a commercially available finite difference time domain (FDTD). The sensitivity to several parameters (LW, LH and SWA) was studied. Both large and small perturbations of these parameters were applied to features with baseline design rules. The OIRs of the differential TSOM images were calculated as the difference between the perturbed state and the baseline state and compared to a noise threshold (OIR = 1) to predict the minimum sensitivity of TSOM to changes in the parameters. In this report, we focus on the small perturbations to determine the minimum threshold for TSOM sensitivity to the specified parameter. The results summarized in Fig. 2(b) show that TSOM has the ability to detect sub-nanometer differences in the dimensions.

3. Conclusion

In summary, experimental work demonstrates sub-nanometer dimensional sensitivity using the TSOM method, even with visible wavelengths. The simulations demonstrate the potential usefulness of TSOM for 3D shape process monitoring of lines as small as 16 nm. Sensitivity to various perturbations appears sufficient and distinct. In addition to economical metrology hardware, the substantially smaller area targets needed result in further potential cost savings. If paired with scanning electron microscopy, scatterometry (optical critical dimension) or other widely-used tools, an inexpensive tool such as TSOM could be a key component of a good hybrid metrology solution.

4. References

[1] R. Attota and R. G. Dixson, "Resolving three-dimensional shape of sub-50 nm wide lines with nanometer-scale sensitivity using conventional optical microscopes," Appl. Phys. Lett. 105(2014).

[2] M. V. Ryabko, S. N. Koptyaev, A. V. Shcherbakov, A. D. Lantsov, and S. Y. Oh, "Method for optical inspection of nanoscale objects based upon analysis of their defocused images and features of its practical implementation," Opt. Express 21, 24483-24489 (2013).

[3] R. K. Attota and H. Kang, "Parameter optimization for through-focus scanning optical microscopy," Opt. Express 24, 14915-14924 (2016).

[4] R. Attota, "Noise analysis for through-focus scanning optical microscopy," Opt. Lett. 41, 745-748 (2016).

[5] R. K. Attota and H. Park, "Optical microscope illumination analysis using through-focus scanning optical microscopy," Opt. Lett. 42, 2306-2309 (2017).

[6] R. Attota, B. Bunday, and V. Vartanian, "Critical dimension metrology by through-focus scanning optical microscopy beyond the 22 nm node," Appl Phys Lett 102(2013).

Non-tunneling origin of the 1/f noise in SiC MOSFET

Kin P Cheung and Jason P Campbell
National Institute of Standards & technology, Gaithersburg, MD, USA

Abstract: **It has long been established that MOSFET random telegraph noise and the cumulative 1/f noise is the result of inversion charge tunneling in and out of bulk traps in the gate oxide near the interface. The tunneling nature is a key concept upon which the technique of trap profiling using 1/f noise is based on. In this work, we examine the tunneling pathways in SiC MOSFETs and show that the 1/f noise observed in SiC MOSFETs is likely of a completely different nature involving above band edge interface states that do not require tunneling to capture the inversion charge.**

While the origin of 1/f noise remains a challenging subject, a consensus had been reached for the case of silicon MOSFETs [1-5]. It is generally accepted that random telegraph noise (RTN) and its cumulative effect, 1/f noise, in silicon MOSFETs is due to capture and emission of charges in trap states in the gate dielectric. It is universally stated that the capture and emission of charges by these states are by a tunneling mechanism which can only happen to bulk traps near the interface. Indeed, this tunneling nature is the foundation of the technique of using 1/f noise for trap profiling [6, 7]. In this work, we examine the available tunneling pathways to show that the RTN and 1/f noise in SiC MOSFETs is not due to bulk traps and does not have a tunneling origin.

MOSFET RTN is generally accepted as the result of charge capture/emission by single bulk traps and 1/f noise is the superposition of RTN events from many traps with different time constants (capture and emission) [8-11]. To explain the time constants as well as the observed thermal activation of RTN, it was concluded that the tunneling process must be inelastic in nature and that a lattice relaxation must be involved in the charge capture process. Working backwards from the measured time constants, the estimated lattice relaxation was between 20 meV to 150 meV [12, 13]. This small relaxation energy upon the capture of a charge is remarkable in that it is not consistent with the flexible network nature of amorphous SiO_2 [14]. Furthermore, direct measurements of SiO_2 defect relaxation energy found a value of 1.5 eV [15, 16]. Since the small relaxation energy was obtained by working backwards from measured time constant, this discrepancy clearly points to basic conflict in the most important parameter in the conventional RTN model. In fact, if one accepts the 1.5 eV relaxation energy, RTN cannot be explained by the existing model.

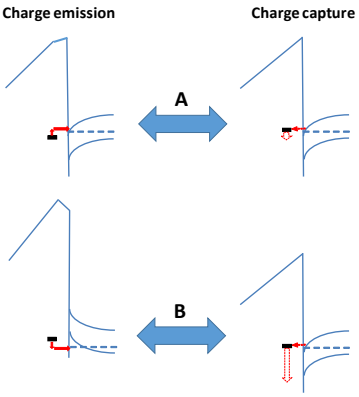


Fig. 1 RTN models showing charge capture and emission through tunneling. A represents the conventional model involving the "sheet charge approximation" and B represents the new model involving the local field of individual trapped charge.

In a recent paper [14], it was shown that the smaller relaxation energy of the existing model originates from the use of the universally accepted "sheet charge approximation" in treating the problem of trapped charge. Instead of this approximation, the recent work aims to capture the localized nature of each trapped charge and its associated electric field [14]. This approach has led to a new understanding of RTN that is compatible with the 1.5 eV relaxation. Fig. 1 illustrates the conventional and the new [14] RTN model of charge capture/emission.

Notice that fig. 1B indicates that the local field of a trapped charge changed the band bending from inversion to strong accumulation after a charge is captured. Notice also the large difference in energy relaxation in the charge capture process between the two models as indicated by the broken red arrow pointing downward. This is the most important difference. as in the conventional model, emission proceeds back to the conduction band where charge capture originated. In the new local field model, emission proceeds to the valence band instead. We emphasize that the local field as well as the strong

Campbell, Jason; Cheung, Kin.
"Non-tunneling origin of the 1/f noise in SiC MOSFET."
Paper presented at 2018 International Conference on IC Design and Technology, Otranto, Italy. June 4, 2018 - June 6, 2018.

relaxation are physical realities that have been unheeded in the conventional model, not approximations introduced to simplify the treatment. On the other hand, the "sheet charge approximation" was introduced to simplify the treatment and the small relaxation energy was a fitting parameter under this approximation. Thus, we argue that the new model is a big step towards a more complete picture of charger capture/emission.

Notice that the 1.5 eV relaxation is larger than the silicon band gap. This has important implications to silicon devices and conveniently locates the electron-captured relaxed trap at an energy favorable to for reemission to the valence band. Although the emission is no longer to the conduction band, the tunneling nature of the RTN is preserved. When dealing with wide band gap semiconductor devices such as SiC MOSFETs, this capture/emission pathway is no longer possible. Yet, it was reported that SiC MOSFETs have 1/f noise behavior very similar to those of Si MOSFETs [17]. This problem is illustrated in fig. 2 for SiC/SiO_2 MOSFETs.

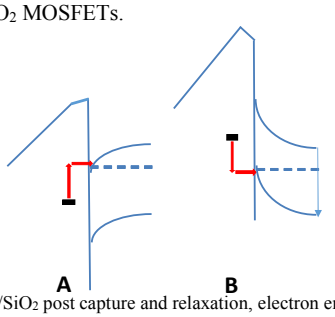


Fig. 2 SiC/SiO_2 post capture and relaxation, electron emission pathway for A, the conventional model involving "sheet charge approximation" and B, new model involving local field.

Since the 1.5 eV relaxation is an experimentally measured value [15,16] that is consistent with the physics of a flexible SiO_2 network [14], it must be accounted for in both the conventional model and the new model. Fig. 2A shows, for the conventional model, that the emission process must provide at least 1.5 eV of activation energy to occur. The probability to do so is extremely low and therefore would not regularly occur. For the new model shown in fig. 2B, the situation is also problematic. Since the bandgap of SiC is greater than 3 eV, it is easier to reemit to the conduction band rather than to the valence band as shown. When emitting to the conduction band, the activation energy is still 1.5 eV and unfavorable. Thus, we have a situation that both the new and the conventional model fail to explain the RTN and therefore the 1/f noise phenomenon in wide bandgap SiC/SiO_2 MOSFETs.

In [14], it was further discussed that energy relaxation is roughly half if the trapped charge is precisely at the interface. Physically, this is because only half of the media surrounding the charge is polar and flexible, the other half is non-polar and rigid. As the trap location moves further into the bulk of the oxide, the relaxation energy rapidly transitions to the full relaxation value (1.5 eV) within a fraction of a nanometer. Fig. 3 illustrates the energetics of emission for true interface states, very near interface states, and bulk states.

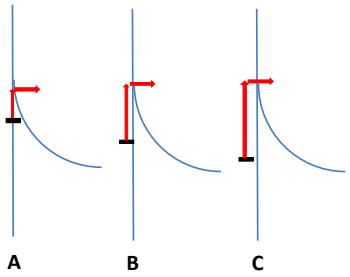


Fig. 3 Emission pathway of trapped electron for three trap locations: A, precisely at the interface; B, very near the interface; and C, near interface.

As shown in Fig. 3A, when the trap is right at the interface, only half of the SiO_2 media can rearrange screen the trapped charge and energy relaxation is essentially cut in half (0.75 eV). When the trapped charge is deeper into the bulk of the oxide, the energy relaxation is 1.5 eV (fig. 3C). In between these two extremes, the relaxation energy varies [14] (fig. 3B). Fig. 3 also depicts a very strong band bending, which is a characteristic of wide band gap semiconductors biased in strong accumulation as would be true in the local field picture [14]. As shown, the activation energy for emission is slightly lower than the relaxation energy for the true interface states near the top of the band, the barrier becomes very thin and tunneling can be more efficient than absorbing more phonons.

For traps located deeper into the oxide, the relaxation energy increases rapidly and the activation energy must reach closer to the top of the conduction barrier

Campbell, Jason; Cheung, Kin.
"Non-tunneling origin of the 1/f noise in SiC MOSFET."
Paper presented at 2018 International Conference on IC Design and Technology, Otranto, Italy. June 4, 2018 - June 6, 2018.

because there is an additional, much higher barrier layer due to the oxide. This is the reason activation energy in fig. 3B rises so dramatically. As the trap gets very deep into the bulk, (fig. 3C), the activation energy equals the relaxation energy. Thus, as soon as the trap is just off the interface, the activation energy makes emission very unlikely. We therefore conclude that only the interface states have the potential to explain the RTN and therefore RTN and 1/f noise. Tunneling is not involved, at least for the capturing process.

As pointed out by McWhorter [8], another condition for 1/f noise is that the capture and emission must involve many traps with a broad range of time constants. The emission time of an electron in a relaxed interface state is:

$$\tau = \frac{1}{\nu e^{(-\Delta E/kT)}} \qquad (1)$$

where ν is the lattice vibration frequency taken as 10^{13}/s; ΔE is the activation energy; k is Boltzmann's constant; T is temperature. Fig. 4 shows the emission time and activation energy relationship.

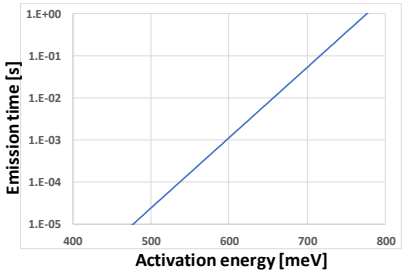


Fig. 4 Emission time as a function of activation energy.

For experimentally observable noise, the time constant ranges from tens of microseconds to seconds. In other words, the activation energy must be between 500 meV and 800 meV. Such a range is possible if there is a high density of interface traps at energy around the conduction band edge. Fig. 5 shows how a range of near-band edge trap energy produces a range of activation energies.

For observable RTN and 1/f noise, the capture time must also be in a similar range. Notice that in fig. 5 some of the traps are above the Fermi level, requiring thermal activation to reach. That will reduce the capture probability, or equivalently increase capture time. In addition, when an electron is captured from the inversion layer into one of the interface trap states, it can either reemit right back out or be trapped, meaning relaxation occurs before reemission. At energies above the Fermi level, reemission dominates because relaxation involves an energy barrier corresponding to the configurational difference between the initial empty state without an electron and the final stable state with a captured electron [14]. This means that states above the Fermi level have longer capture time. The combined thermal activation energy to reach these states and the activation energy to reach the new configuration quickly limit the range of energies above the Fermi level that can be involved. The boundary can be obtained by realizing that the emission size limited the top end of energy barrier to be 500 meV below the conduction band edge. Assuming 750 meV of energy relaxation, the maximum trap energy before capture is 250 meV above the band edge. Due to quantum confinement, the Fermi level is somewhere within the first sub-band which is 100 meV to more than 200 meV above the conduction band edge. We therefore only need to thermally access a few tens of meV, or a couple of kTs.

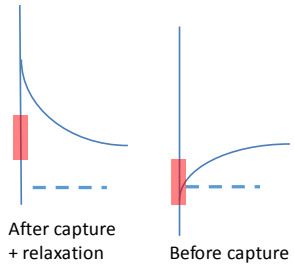


Fig. 5 High density of interface traps at energy around the conduction band edge can produce a range of activation energy after electron capture and relaxation.

For interface states with energy below the Fermi level, capture becomes more efficient because reemission requires energy. The competition between capture (with relaxation) and reemission increasingly favors capture. At some energy below the Fermi energy, capture becomes too efficient and the trap will always be filled. At this point noise cannot be observed. While we do not know the size of the energy barrier involved in configuration change, we can get a lower bound of the interface state energy from the emission consideration. With the maximum activation energy for emission at 800 meV and the relaxation energy at

Campbell, Jason; Cheung, Kin.
"Non-tunneling origin of the 1/f noise in SiC MOSFET."
Paper presented at 2018 International Conference on IC Design and Technology, Otranto, Italy. June 4, 2018 - June 6, 2018.

750 meV, we know that the lowest energy interface states that contributes to RTN and 1/f noise is at most 50 meV below the conduction band edge.

High quality silicon MOSFETs do not have high densities of interface states near the band edge [18] so the tunneling nature of noise is seemingly correct. For SiC MOSFETs, it is well known that high density of interfaces states exists at the band edge [18, 19]. Since we already discussed that the tunneling picture does not fit wide bandgap MOSFETs, this band edge interface states based, non-tunneling mechanism has solid foundation. We note that 1/f noise in SiC MOSFETs has been explained previously using interface states [21], but the model employed was for the tunneling process and therefore problematic.

We note that the exact values of relaxation are somewhat approximate as the defect identity does play some role in the overall relaxation energy. In the case of SiC MOSFETs, where substantial doses of nitrogen are incorporated at the interface and near interface oxide regions, the relaxation energies may differ slightly than those reported here for the pure SiO_2 network. The addition of nitrogen into the SiO_2 glass network will 'stiffen' the resultant insulator network and reduce the relaxation energies somewhat. However, even high-density nitrogen incorporation would leave the average SiO_2 network (average bond energies) essentially unchanged. Consequently, we take the pure-SiO_2 relaxation energies discussed above as very good approximations.

Summary: As a flexible network of highly polar bonds, SiO_2 produces a large lattice relaxation energy of 1.5 eV. When this physical reality is accounted for, RTN and therefore 1/f noise cannot be explained in the conventional model where the "sheet charge approximation" is used to treat trapped charges. Abandoning this approximation and recognizing the localized nature of the trapped charge and its associated electric field, RTN of silicon MOSFETs can be explained without abandoning the 1.5 eV relaxation energy. However, when wide band gap semiconductors such as SiC MOSFETs are involved, the tunneling based models, conventional or local field, involving bulk traps do not work. Acknowledging that there are high densities of interface states at the band edge in SiC MOSFETs, a new, non-tunneling based model involving true interface states is developed that can successfully explain the observed RTN and consequent 1/f noise in SiC MOSFETs.

[1] Dutta, P. and P. M. Horn, "Low-frequency fluctuations in solids: 1/f noise." Rev. Mod. Phys. **53**(3): 497-516(1981).
[2] Weissman, M. B. "1/f noise and other slow, nonexponential kinetics in condensed matter." Rev. Mod. Phys. **60**(2), 537-571(1988).
[3] van der Ziel, A. "Unified presentation of 1/f noise in electron devices: fundamental 1/f noise sources." Proc. IEEE **76**(3), 233-258(1988).
[4] Ghibaudo, G. "Unified formulation of trapping noise and fluctuation in MOS devices." NOISE IN PHYSICAL SYSTEMS, Sep 21 - 25 1989, Budapest, Hungary, Publ by Akad Kiado.
[5] Hung, K. K., P. K. Ko, et al. "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors." IEEE Trans. Electron Dev., **37**(3), 654-665(1990).
[6] Celik-Butler, Z. and T. Y. Hsiang, "Determination of Si-SiO2 interface trap density by 1/f noise measurements." IEEE Trans. Electron Dev., **35**(10), 1651-1655(1988).
[7] Jayaraman, R. and C. G. Sodini, "A 1/f noise technique to extract the oxide trap density near the conduction band edge of silicon." IEEE Trans. Electron Dev., **36**(9), 1773-1782(1989).
[8] McWhorter, A. L. "1/f noise and related surface effects in germanium." (1955). Ph.D. thesis, Dept. of Electrical Engineering. Boston, Massachusetts Institute of Technology.
[9] Ralls, K. S., W. J. Skocpol, et al. "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/f) Noise." Phys. Rev. Lett. **52**(3), 228-231(1984).
[10] Howard, R. E., W. J. Skocpol, et al. "Single electron switching events in nanometer-scale Si MOSFET's." IEEE Trans. Electron Dev., **32**(9), 1669-1674(1985).
[11] Uren, M. J., D. J. Day, et al. "1/f and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors." Appl. Phys. Lett. **47**(11), 1195-1197(1985).
[12] M. J. Kirton and M. J. Uren, "Capture and emission kinetics of individual Si:SiO2 interface states," Appl. Phys. Lett., **48**(19), 1270–1272(1986).
[13] A. Palma, A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller, and J. A. López-Villanueva, "Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide–semiconductor structures," Phys. Rev. B, Condens. Matter, **56**(15), 9565–9574(1997).
[14] Cheung, K. P., D. Veksler, et al. "Local Field Effect on Charge-Capture/Emission Dynamics." IEEE Trans. Electron Dev. **64**(12), 5099-5106(2017).
[15] Y. Lu and C.-T. Sah, "Thermal emission of trapped holes in thin SiO_2 films," J. Appl. Phys., **78**(5), 3156–3159(1995).
[16] S. Takagi, N. Yasuda, and A. Toriumi, "Experimental evidence of inelastic tunneling and new I-V model for stress-induced leakage current," Int. Electron Devices Meet., Washington, DC, USA, Dec. 1996, pp. 323–326.
[17] Rumyantsev, S. L., M. S. Shur, et al. "Si-like low-frequency noise characteristics of 4H-SiC MOSFETs." Semiconductor Science and Technology **26**(8), 1-5(2011).
[18] Ryan, J. T., R. G. Southwick, et al. "On the "u-shaped" continuum of band edge states at the Si/SiO2 interface." Appl. Phys. Lett., **99**(22). 223516(2011).
[19] Nakazawa, S., T. Okuda, et al. "Interface Properties of 4H-SiC (1120) and (1100) MOS Structures Annealed in NO." IEEE Trans. Electron Dev., **62**(2), 309-315(2015).
[20] Dhar, S., S. Haney, et al. "Inversion layer carrier concentration and mobility in 4H-SiC metal-oxide-semiconductor field-effect transistors." J. Appl. Phys. **108**(5), 054509(2010).
[21] Zhang, C. X., E. X. Zhang, et al. "Origins of Low-Frequency Noise and Interface Traps in 4H-SiC MOSFETs." IEEE Electron Dev. Lett. **34**(1): 117-119(2013).

Campbell, Jason; Cheung, Kin.
"Non-tunneling origin of the 1/f noise in SiC MOSFET."
Paper presented at 2018 International Conference on IC Design and Technology, Otranto, Italy. June 4, 2018 - June 6, 2018.

Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances

Jodie G. Pope[a], Keith A. Gillis[a], Michael R Moldover[a], James B. Mehl[b], and Eric Harman[c].
[a] *Fluid Metrology Group, NIST, Gathersburg, MD 20899, USA*
[b] *36 Zunuqua Trail, P.O. Box 307, Orcas, WA 98280-0307, USA*
[c] *Colorado Engineering Experiment Station, Inc (CEESI), USA*

Abstract

We characterized a 1.8 m^3, nearly-spherical, steel shell at pressures up to 7 MPa for use as a gas flow standard. For pressure, volume, temperature, and time measurements, the shell's cavity will collect gas; for blow-down measurements, the shell will be a gas source. We measured the cavity's microwave resonance frequencies f_{micro} to determine its pressure- and temperature-dependent volume: $V_{micro}(P, T) = 1.84740\ m^3 \times [1 + \alpha(T–295\ K) + \kappa P]$ with a fractional uncertainty of 0.011 % at a 68 % confidence level. The coefficients α and κ were consistent with the dimensions and properties of the steel shell. The microwave-determined volume V_{micro} was consistent, within combined uncertainties, with V_{gas} the volume determined by a gas expansion method: $V_{micro}/V_{gas} - 1 = (2 \pm 14)\times10^{-5}$. When the shell was filled with gas, measurements of its acoustic resonance frequencies f_{acoust} and of the pressure quickly and accurately determined the mass of the gas in the shell, even when temperature gradients persisted. [K. A. Gillis *et al*. *Metrologia*, **52**, 337 (2015)] After raising the nitrogen pressure in the shell from 0.1 MPa to 7.0 MPa in 45 minutes, the top of the shell was ≈20 °C warmer than the bottom of the shell. Despite this large thermal gradient, the mass M_{acoust} of gas determined from acoustic resonance frequencies settled to within 0.01 % of its final value after 5 h. Following a smaller pressure change of 0.3 MPa, the top-to-bottom temperature difference was 1.5 °C and M_{acoust} settled to its final value in just 0.5 h.

Nomenclature

a	Inner radius
α	Thermal expansion coefficient
β_a	Acoustic virial coefficients
B	Density virial coefficients
c_0	Speed of light in vacuum
D_T	Thermal diffusivity of gas in BBB
f_{micro}	Microwave resonance frequency
$\left\langle f_{ln}^{\sigma} \right\rangle$	Average frequency of a microwave multiplet.
f_{acoust}	Acoustic resonance frequency
γ_0	$\equiv C_p/C_V$, zero-density ratio of the constant-volume specific heat to the constant-pressure specific heat.
j_0	Zeroth order spherical Bessel function
κ	Pressure expansion coefficient
M_{acoust}	Mass determined from f_{acoust}
M_{BBB}	Mass of gas in the BBB
M_∞	Equilibrium mass in BBB at time infinity
M_{grav}	Mass of gas determined by a weighing technique
n_g	Refractive index of a gas
P	Pressure

1. Introduction

During 2016, the value of natural gas metered in pipelines in the United States was approximately $90 billion. To ensure equity at each transfer of custody, accurate metering is required, both in the US and international markets. NIST calibrates natural gas flow meters and has an ongoing research program to improve the accuracy these calibrations [1]. At present, NIST traces the calibration of pipeline-scale natural gas flowmeters to NIST's primary gas flow standard that uses the pressure, volume, temperature, and time (*PVTt*) technique [2]. This primary standard relies on a well-characterized, carefully-thermostated [$u(T)$ = 6 mK][1], 0.67 m^3 collection vessel that operates at pressures up to 0.15 MPa. The primary standard is used to calibrate, one at a time, 21 critical flow venturis (CFVs), each 5.2 mm in diameter. The 21 CFVs are used in parallel to calibrate several 25 mm-diameter CFVs, one at a time [3]. This use of 21 CFVs in parallel is the first of 6 stages of scale-up that uses both CFVs and

[1] Unless otherwise stated, all uncertainties are one standard uncertainty corresponding to 68 % confidence level.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Nomenclature (continued)

ρ	Gas density
$\varphi_i(\mathrm{r})$	Acoustic velocity potential of a standing wave as a function of radius in a sphere.
T	Temperature
$\langle T\rangle_{\mathrm{sh}}$	Average shell thermometer readings
$\langle T\rangle_V$	Volume average temperature
$\langle T\rangle_\varphi$	Acoustic mode-dependent average temperature
τ	Time required for a changing quantity to reach 63.2 % of its new value.
u	Standard uncertainty
u_r	Relative standard uncertainty such that $u_\mathrm{r}(x) = u(x)/x$
V_{micro}	Volume of BBB determined by microwave resonance frequencies
V_{acoust}	Volume of BBB determined by acoustic resonance frequencies
V_{gas}	Volume of BBB determined by gas expansion
V_{addenda}	Volume of ports and fittings on BBB that are not detected by the long-wavelength microwaves and sound waves.
$V_{\mathrm{inventory}}$	Volume of the connecting plumbing between the BBB and the reference volume for the gas expansion measurement.
V_{ref}	Volume of the reference used in the gas expansion measurement.
w	Speed of sound
$\langle w\rangle$	Average speed of sound
x_w	Mole fraction of water in a gas
ξ^{σ}_{ln}	Exactly-known microwave eigenvalue of a spherical cavity, σ = TM or TE
Z	$= P/(\rho RT)$, compressibility factor.
z_{acoust}	Exactly-known acoustic eigenvalue of a spherical cavity

turbine meters. The scale-up begins with flows of ≈10 g/s and ends with flows encountered in large natural gas pipelines: ≈500 kg/s at pressures near 7 MPa. Each stage of scale-up adds cost and uncertainty to the calibration of large meters. The purpose of this work is to reduce the number of stages in the traceability chain by starting with a novel, primary gas-flow standard that operates at higher pressures and flow rates than NIST's present primary standard.

In this paper, we characterize a large volume (1.8 m^3) enclosed by a nearly-spherical, steel shell. This volume may be used at pressures up to 7 MPa as either a gas source or as a gas collector during calibrations of CFVs and other meters. Informally, we call the steel shell the Big Blue Ball (BBB). We used microwave resonance frequencies f_{micro} to determine the $V_{\mathrm{micro}}(P, T)$ of the BBB, where V, T, P are the volume, temperature and pressure, respectively. A summary of the microwave results is:

$$\begin{aligned} V_{\mathrm{micro}}(P,T) &= 1.84740\ \mathrm{m}^3(1\pm 1.1\times 10^{-4})[1+\alpha(T-295\mathrm{K})+\kappa P), \\ \alpha/\mathrm{K} &= 5.24\times 10^{-5}(1\pm 0.081), \\ \kappa/\mathrm{MPa} &= 1.790\times 10^{-4}(1\pm 0.013). \end{aligned} \quad (1)$$

Figure 1 displays $V_{\mathrm{micro}}(P, T)$ and the deviations from Eq. (1). We also used a gas expansion technique to independently measure the volume of the BBB, V_{gas}(0.1 MPa, 295 K). As shown in Fig. 1(a), the two techniques agreed within combined uncertainties: $V_{\mathrm{micro}}/V_{\mathrm{gas}} - 1 = (2 \pm 14)\times 10^{-5}$.

In addition to measuring $V_{\mathrm{micro}}(P, T)$ we determined M_{BBB}, the mass of the gas in the BBB, by measuring acoustic resonance frequencies f_{acoust} and the pressure. First, we discuss determining M_{BBB} when the BBB and the gas within it are in thermal equilibrium; then we discuss the effects of temperature gradients.

We determined the speed of sound $w(P, T)$ in the gas using the relation

$$w = f_{\mathrm{acoust}}\left(6\pi^2 V_{\mathrm{micro}}\right)^{1/3} / z_{\mathrm{acoust}}\ , \quad (2)$$

where z_{acoust} is an exactly-known eigenvalue of a spherical cavity [4]. With w known, we determined the gas density $\rho(P, w)$ using tables that we generated using the computer package REFPROP [5]. Finally, we determined the mass of the gas in the BBB from the product

$$M_{\mathrm{acoust}} = V_{\mathrm{micro}}\rho(P,w)\,. \quad (3)$$

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

To provide insight into the sensitivity of M_{acoust} to our measurements and to the REFPROP-generated tables, we follow Gillis *et al*. [6] in using the exact (but truncated) virial expansions of $P(\rho, T)$ and $w(\rho, T)$ to calculate $\rho(P, w)$. The result is:

$$M_{\text{acoust}} = \frac{\gamma_0 P V_{\text{micro}}}{w^2}\left[1 + (\beta_a - B)\frac{P}{RT} + \ldots\right] . \qquad (4)$$

In Eq. (4), $\gamma_0 \equiv C_P/C_V$ is the zero-density ratio of the constant-pressure specific heat to the constant-volume specific heat, and β_a and B are the acoustic and the density virial coefficients, respectively. In Eq. (4), the real-gas term, $(\beta_a - B)P/(RT)$ is 0.073 for nitrogen at 295 K and 7 MPa, and the sum of all the terms in the brackets is 1.097. The uncertainty of M_{acoust} is the combined uncertainties of the measured quantity $\gamma_0 P V_{\text{micro}}/w^2$ and the quantity in the brackets obtained from the equation of state. For nitrogen, argon, methane, and several other gases, the relative uncertainty of $u_r(\rho(P, w)) \le 3\times10^{-4}$ for $P \le 7$ MPa and $T \approx 20$ °C [5].

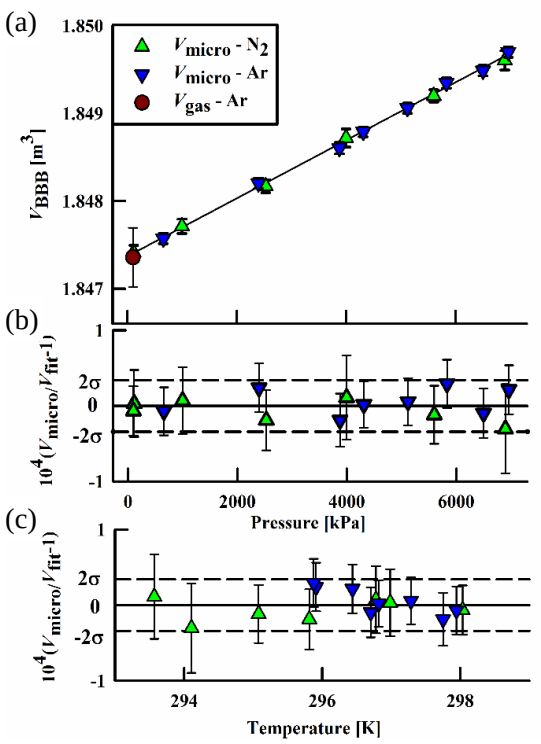


Figure 1. (a) Effect of pressure on the volume of the BBB at 295 K measured using microwaves and gas expansion. (b) fractional volume deviations from Eq. (1) as a function of pressure at 295 K, and (c) as a function of temperature at 0 MPa. The errors bars are one standard deviation of the measurements made with 3 microwave modes. The fractional standard deviation of the data from the fit was $\sigma = 17\times10^{-6}$.

Pumping gas into the BBB or bleeding gas out of the BBB is accompanied by flow work that generates temperature gradients that can be much larger than the gradients generated by fluctuations in the ambient temperature. In one test, the shell was insulated by covering it first with a layer of 2 cm-thick air-filled bubble wrap and then with a 5 mm-thick double reflective, air-filled wrap. The nitrogen pressure in the shell was raised from 0.1 MPa to 7.0 MPa in 45 minutes. Then, the top of the shell was ≈20 °C warmer than the bottom of the shell, as determined by external thermometers, as shown in Fig. 2(a). Subsequently, M_{BBB}, as computed from the average shell thermometer readings $\langle T\rangle_{\text{sh}}$ using $M_{\text{BBB}} = PV_{\text{BBB}}/(Z\langle T\rangle_{\text{sh}})$, decayed towards its final value M_∞ with a time constant $\tau \approx 5$ h, shown by the orange circles in Fig. 2(b). Here, $Z = P/(\rho RT)$ is the compressibility factor from REFPROP [5]. For the same test, M_{acoust} from Eq. (4), approached its final value with a time constant $\tau \approx 1.5$ h (the blue triangles in Fig. 2(b)). In a second, more-realistic test, the shell was not insulated and the pressure in the shell was raised from 6.6 MPa to 6.9 MPa. In this case, $PV_{\text{BBB}}/(Z\langle T\rangle_{\text{sh}})$ approached equilibrium with $\tau \approx 1.8$ h; for M_{acoust}, $\tau \approx 0.8$ h. The more-rapid equilibration of M_{acoust} is a distinct advantage of the acoustic method of determining M_{BBB}.

In deriving Eq. (3), we assumed that the gas was isothermal or, equivalently, $\rho(P, w)$ was constant throughout the cavity. When temperature gradients are present, $\rho(P, w)$ is a function of position within the cavity. Then, for Eqs. (3) and (4) to remain valid, we must replace ρ with $\langle\rho\rangle$ and $1/w^2$ with $\langle 1/w^2\rangle$, where the angled brackets indicate averages over the cavity's volume. In Section 4.2 below, we argue that $1/w^2$, as determined by Eq. (2) is a good

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

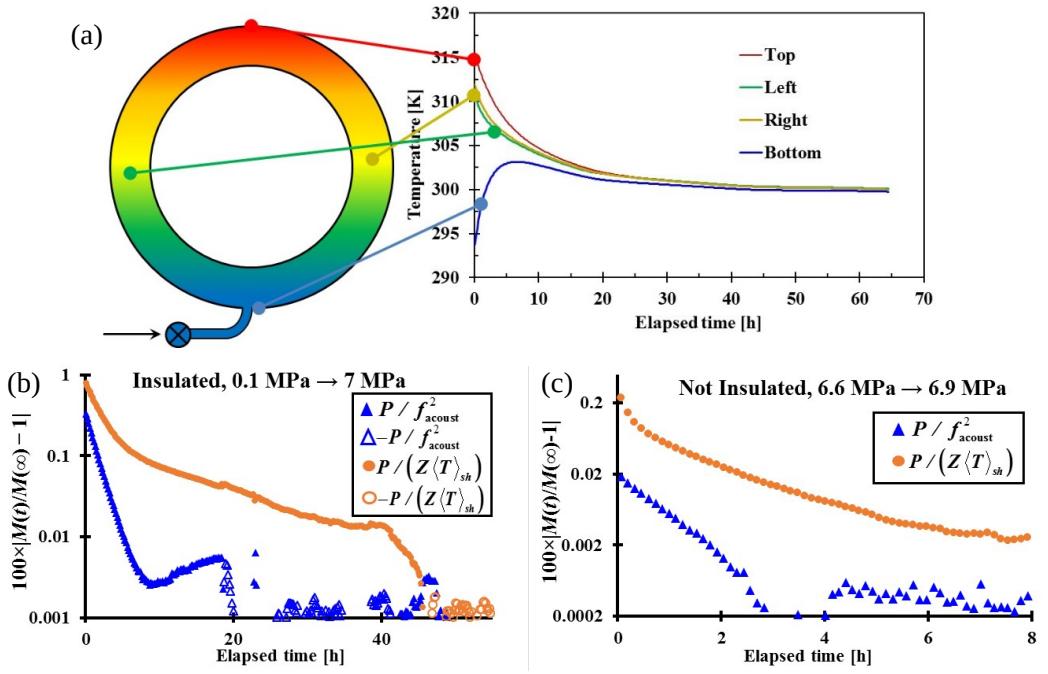


Figure 2. (a) Temperature differences on the outside of the BBB following pressurizing from 100 kPa to 7 MPa. (b) Approach to equilibrium of P / f_{acoust}^2 and $P / \left(Z \langle T \rangle_{sh}\right)$, where $\langle T \rangle_{sh}$ is the average of the 4 temperature sensors shown in (a). (c) Similar to (b) following a smaller pressure change and without insulation surrounding the BBB. Note: negative values of the ordinate are plotted as open symbols.

approximation of $\langle 1/ w^2 \rangle$. Indeed, for plausible temperature distributions, the error of this approximation is on the order of $(\Delta T/T)^2$, where ΔT is a typical temperature difference in the cavity. (For ΔT = 3 K and T = 300 K, the fractional error is of order 10^{-4}).

By using the BBB as a gas collection vessel, [2] we have scaled up our experience in characterizing collection vessels by a factor of almost 10 in pressure, from 0.6 MPa to 7 MPa, and by a factor of 6 in volume, from 300 liters to 1800 liters [6]. The present work is an interim step towards gas-flow measurements traced to standards of frequency and pressure, and to the literature of $\rho(P, w)$ data. This contrasts with NIST's existing gas-flow measurements traced standards of temperature and pressure and to the literature of $Z(P, T)$ data. If this work is successful, it will shorten the scale-up chain and improve equity in large-scale natural gas transactions.

In a future publication, we will compare M_{acoust} with the M_{grav}, the mass of gas in the BBB determined by a weighing (gravimetric) technique. Our preliminary results indicate that $|M_{\text{acoust}}/M_{\text{grav}} - 1| < 5 \times 10^{-4}$.

The remainder of this paper is organized as follows: Section 2 describes the BBB; Section 3 describes how the volume was determined; Section 4 explains the advantages of using acoustic resonance frequencies versus thermistors for temperature measurements and describes how the average temperature is determined using radial acoustic modes; and Section 5 is the discussion.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

(a) (b)

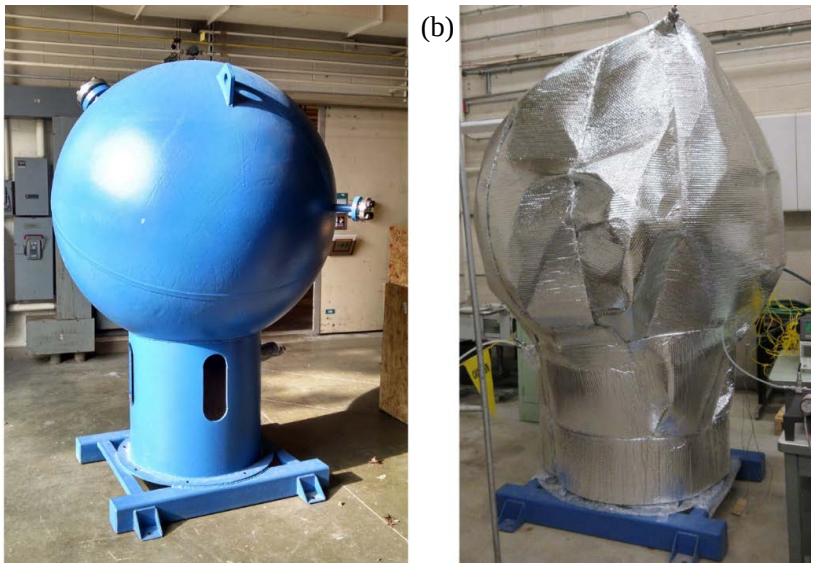


Figure 3. The BBB uninsulated (a) and insulated (b) for gas expansion experiments.

2. BBB description

The BBB was a 1.5 m diameter, nearly-spherical shell made of carbon steel of unknown composition. It had a nominal mass of 1222 kg including a welded metal pallet. See Fig. 3(a). In Fig. 3(b), the BBB is wrapped with an insulating blanket that was used during gas expansion measurements (Section 3). Table 1 lists the properties of the BBB. We believe that the BBB was manufactured in the early 1970's. Before NIST acquired the BBB, it had been stored outdoors for years. During storage, the shell's ports were open; at times the shell contained rainwater. We prepared the BBB by sandblasting and painting the exterior and by hydrostatically testing it to 9.7 MPa. To clean the BBB's interior, we poured methanol through the ports and removed it through the drain pipe. This rinsing was repeated until the drained liquid ran clear. Then, we dried the BBB by evacuating it (< 100 Pa) for 4 days. We determined the BBB's pressure-dependent internal volume while it was filled with argon. As the argon was removed from the BBB, we measured its dew point. The dew point corresponded to a water mole fraction $x_w < 4\times10^{-5}$, which had negligible effects[2] on our measurements. When the BBB was initially sealed for gas expansion experiments, the fractional leak rate, as determined from acoustic measurements [7], was approximately $(1/M_{BBB})(dM_{BBB}/dt) = -4.9\times10^{-5}$/h; we corrected our data for this leak.

The BBB had 5 small volumes (addenda) that were always connected to the spherical cavity. The addenda included 3 flanged ports and short tubes connecting each port to its flange. The addenda also included the plumbing components that permanently connected a pressure sensor to the un-flanged port. The drain tube led to a valve that was always closed. Table 2 lists the volume associated with each port, the drain, and the pressure sensor. The total volume of these addenda $V_{addenda} = 965\ cm^3 = 5.2\times10^{-4}\ V_{BBB}$. As discussed in Section 3.2, $V_{addenda}$ is not detected by the long-wavelength microwaves and sound waves used in these measurements. Therefore, our comparisons of the gas-expansion and microwave volume determinations (shown in Fig. 1) compare $(V_{gas} - V_{addenda})$ to V_{micro}.

Table 1. Properties of the BBB

Material	carbon steel
Mass	1222 kg
Inner Diameter	1.5 m
Wall Thickness	19 mm
Volume	1.8 m³
Hydrostatic Test	9.7 MPa
Working Pressure	7 MPa
Operating Temperatures	–40 °C to 49 °C

The diameter of the largest port was only 5 cm; therefore, an unaided visual inspection of the inside surfaces

[2]Adding water vapor at $x_w = 4\times10^{-5}$ to argon will increase the square of the speed of sound by the fraction 8.9×10^{-7} and, at 5 MPa, increases the dielectric constant by approximately 2.1×10^{-5}.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

was uninformative. We used an articulating-head borescope to inspect the internal surfaces. To estimate the dimensions of objects in borescope images, we attached a probe (8 mm long and 5 mm wide) to the head of the borescope. Rust and other debris remained inside the BBB after it was rinsed and dried. The images showed the *rough* walls and the circumferential weld, which was approximately 2 cm wide and raised approximately 6 mm from the internal surface. The lip of the weld contained a layer of rust dust of varying thickness that could not be washed away. Despite these defects, we used microwaves to measure the volume of the BBB and used acoustic waves to measure the average gas temperature with sufficient accuracy for the present purposes. We speculate that improved accuracy could be obtained using a cleaner pressure vessel.

Table 2. Volumes of addenda measured by the gas expansion method but not by the microwave method.

Component	Volume [cm^3]	Volume/V_{BBB}
2.5 cm diameter drain tube	409	2.2×10^{-4}
P_{BBB} sensor plumbing	63	3.4×10^{-5}
5 cm diameter port	396	2.1×10^{-4}
2.5 cm diameter port	77	4.2×10^{-5}
1.3 cm diameter port	20	1.1×10^{-5}
TOTAL	965	5.2×10^{-4}

3. Volume determination

3.1 Gas expansion

We measured the internal volume of the BBB by applying the same mass-conserving, volume expansion principle and the same reference volume as reported in Ref. [8]. The result was:

$$V_{gas} = 1.84736\left(1 \pm 9.2\times10^{-5}\right)\ \mathrm{m}^3\ . \tag{5}$$

We expanded argon gas (99.999 % purity) from the BBB into a gravimetrically-calibrated, precisely-thermostated [short term: ±1 mK; $u(T)$ = 6 mK] reference volume located in a laboratory where the temperature varied from 295.3 K to 296.5 K. The final pressures in the BBB after expansions ranged from 35.67 kPa to 93.13 kPa.

Figure 4 is a schematic of the experimental set-up. A 1.27 cm-diameter tube connected the BBB's drain to the reference volume. The internal volume of this tube (1538 cm^3) was designated "$V_{inventory}$". In Fig. 4, the components of $V_{addenda}$ and $V_{inventory}$ are enclosed by a dashed curve; these components were all outside of the thermostated water bath. Initially, the reference and inventory volumes (V_{ref} + $V_{inventory}$) were evacuated to a pressure less than 1 kPa while (V_{BBB} + $V_{addenda}$) was filled with argon to 126 kPa. After thermal equilibration, the initial temperatures and pressures were recorded. Then, the valve on the BBB's drain was opened to expand the argon from (V_{BBB} + $V_{addenda}$) into (V_{ref} + $V_{inventory}$). After equilibration (typically ≈15 minutes), final temperature and pressure measurements were made. The expansion was repeated 4 times, each time by closing the valve on the BBB's drain and evacuating (V_{ref} + $V_{inventory}$) and then opening the valve to expand the argon from (V_{BBB} + $V_{addenda}$) into the evacuated volumes. After 4 expansions, the final pressure was 35 kPa. This sequence of 4 expansions was performed 3 times. A fourth experiment was performed where V_{ref} was pressurized to 128 kPa and $V_{inventory}$ and V_{BBB} evacuated to a pressure < 1 kPa and the expansion was reversed. The mutual agreement of all four sets of expansions is shown in Fig. 6.

In separate measurements, we determined $V_{inventory}$ using the same mass-conserving, gas-expansion principle. The reference volume V_{ref} was filled with argon to 130 kPa while $V_{inventory}$ was evacuated to a pressure less than 1 kPa. The expansion of the argon from V_{ref} into $V_{inventory}$ yielded the result $V_{inventory}$ = (1358 ± 25) cm³, where the uncertainty is one standard uncertainty.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

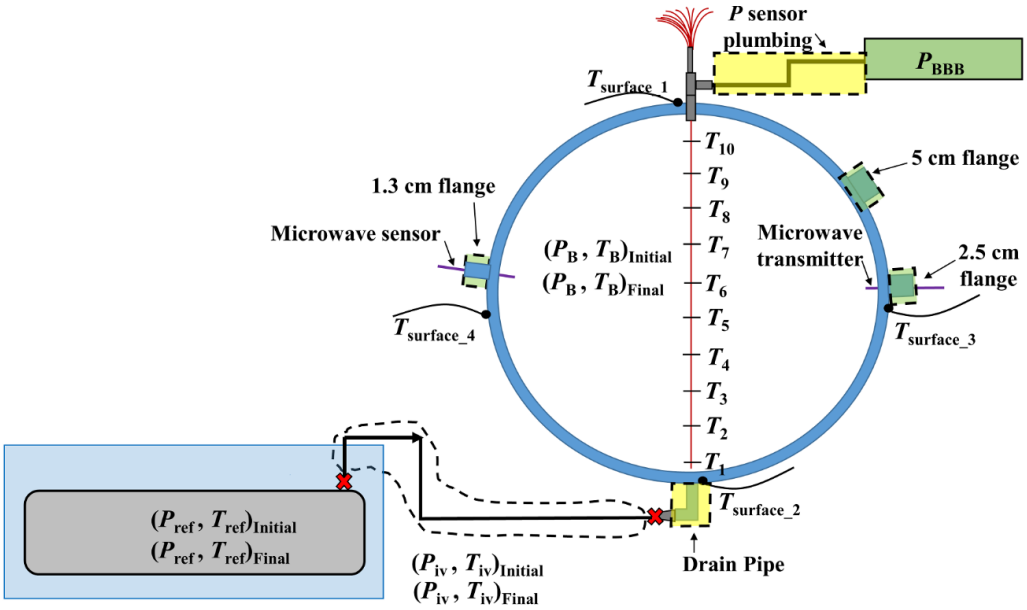


Figure 4. Schematic of BBB gas expansion set-up showing the extra plumbing that the gas expansion method measures but is excluded from the microwave measurements. The subscripts "iv", "ref", and "B" denotes inventory volume, the reference volume and the BBB volume, respectively.

The gas expansion method requires an accurate estimate of the volume-averaged density of the argon in the BBB preceding and following each expansion. This estimate requires a model for the temperature distribution in the argon. The largest uncertainty component of V_{gas} results from the uncertainty of the temperature distribution. Temperature fluctuations in the laboratory, combined with limited air circulation, generated time-dependent temperature gradients in the BBB. (The BBB was too large to fit in the thermostated laboratory with the reference volume.)

Figure 7(b) displays a typical, approximately-linear, vertical, temperature profile in the argon inside the BBB; the top was 0.4 K warmer than the bottom. We determined the volume-averaged temperature[3] using 10 thermocouples, equally-spaced along the vertical diameter of the BBB and by 4 surface thermistors placed equal-distant along the vertical and horizontal axis. (See Fig. 4.) Before insulating the BBB, the temperature uncertainty contributed more than 88 % of the density-change uncertainty. Insulating the BBB decreased the temperature difference between the top and bottom by nearly a factor of 2 and reduced the relative uncertainty of V_{gas} from 0.030 % to 0.019 % (k = 1).

3.2 Microwave measurements

We made an independent determination of the BBB volume by measuring the frequencies of microwave resonances. This technique is described in more detail in Ref. [8]. To measure the microwave resonance frequencies, we installed transmitting and receiving antennas into two of the ports in the BBB (see Fig. 5). One microwave (MW) antenna was installed into the 1.3 cm-diameter port, as shown Fig. 5(b). The other MW antenna was installed along with a sound source (speaker) into the 5 cm-diameter port, as shown in Fig. 5(c). Electrical connection to each antenna was made using a Bayonet Neill – Concelman (BNC) hermetic feedthrough with a 3/8" NPT fitting

[3] Technically $\langle\rho\rangle$ requires the determination of $\langle 1/T\rangle$ not $1/\langle T\rangle$. However, the fractional difference between the two averages is $(\sigma/\langle T\rangle)^2 \approx 10^{-4}$ to lowest order, where σ is the standard deviation of the temperatures.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

mounted into the port flange. The MW antennas were made from semi-rigid coaxial cable with solid copper shield, polytetrafluoroethylene insulation, and silver-plated, solid copper center conductor. As shown in Figs. 5(b) and 5(c), the shield and insulation of the coaxial cables extended from the electrical feedthroughs along the lengths of ports to the inside surface of the BBB. About 2 cm from the opening into the BBB, the coax passed through a small hole in a coarse mesh copper screen and was secured to it with soft solder. The screen performed 3 functions: 1) it rigidly held the coax centered in the duct, 2) it grounded the shield to the metal shell, and 3) it prevented the penetration of microwaves into the duct, yet it allowed gas and sound waves to freely pass. The bare center conductor extended 11 cm further into the BBB forming a straight antenna that coupled to the electric field. The antenna protruded into the BBB by 14 % of the BBB's radius, the same length-to-radius ratio used in Ref [8]. Following Ref. [8], we estimate the protruding center conductors shifted the microwave resonance frequencies by the fraction 2×10^{-5} or less. The antenna in the 1.3 cm-diameter port had a 1 cm-diameter loop to improve coupling to the magnetic field.

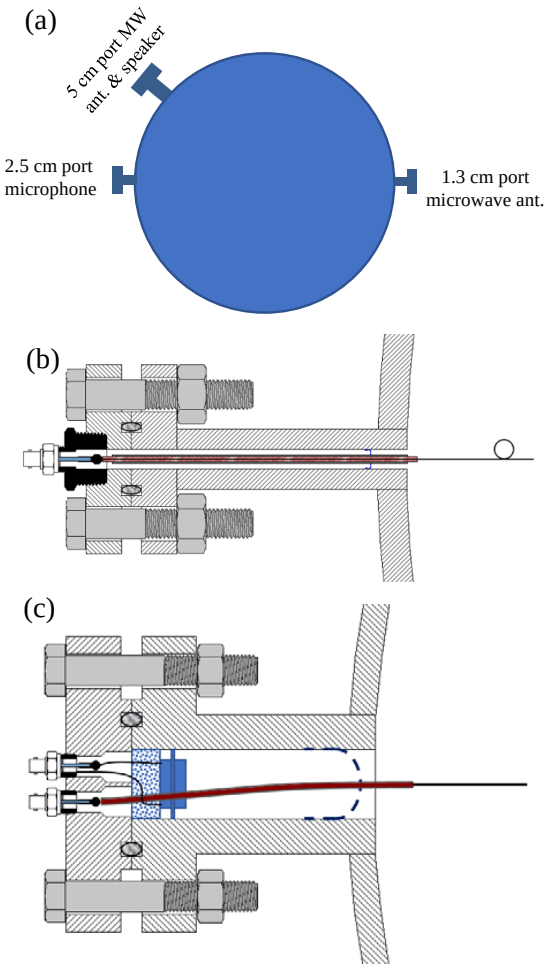


Figure 5. (a) Port locations and uses. (b) Microwave (MW) antenna mounted in the 1.3 cm-diameter port. (c) MW antenna and a 38 mm-diameter loudspeaker (dark blue rectangle) mounted in the 5-cm diameter port. Flexible foam material dampened the sound emitted from the rear of the speaker.

The theory for the electromagnetic modes of a spherical cavity is described in detail in Refs. [9] and [10]. Following conventional derivations in the literature, we divide the modes into two types: transverse electric (TE) and transverse magnetic (TM) modes, according to which field (electric or magnetic, respectively) has its radial component identically equal to zero. Modes are designated TE*ln* and TM*ln*, where *l* and *n* are integers greater than zero that identify the eigenfunction and the resonance frequency f_{ln}^{σ} (with σ = TE or TM). In a perfectly spherical chamber, the resonance frequencies form (2*l*+1)-fold degenerate multiplets; however, the degeneracy may be partially or fully lifted by perturbations in the shape, such as imperfect sphericity or the existence of ports. The resonance frequencies of a gas-filled spherical cavity with radius *a* and a rigid, perfectly conducting wall are given by

$$f_{ln}^{\sigma} = \frac{\xi_{ln}^{\sigma} c_0}{2\pi n_g a} \quad , \tag{6}$$

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

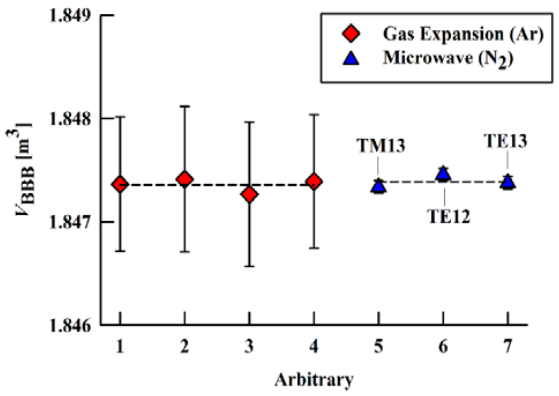


Figure 6. Comparison of the volume of the BBB at 100 kPa and 295 K as determined by gas expansion using argon gas and microwave measurements using nitrogen gas.

where ξ_{ln}^{σ} is an exactly known eigenvalue, c_0 is the speed of light in vacuum, and n_g is the refractive index of the gas. If the resonance frequency is measured and $n_g(T, P)$ is known, then Eq. (6) can be inverted to determine $a(T, P)$. When the degeneracy is lifted by shape perturbations, Eq. (6) still holds to first order if f_{ln}^{σ} is replaced by the average frequency of the multiplet $\left\langle f_{ln}^{\sigma} \right\rangle$ [9]. When the perturbation is small, the multiplet will be partially resolved, and a precise determination of $\left\langle f_{ln}^{\sigma} \right\rangle$ will be difficult if the multiplicity is large. Therefore, we focus on the modes with the lowest multiplicity, the $l = 1$ triplets.

We used a commercially-manufactured, microwave network analyzer to measure the electromagnetic power transmitted from one antenna through the BBB to the other antenna as a function of frequency. The measurements were conducted while the BBB was filled with high purity nitrogen or argon. We measured the complex response (the parameter S12) in the vicinity of 3 modes TM13, TE12, and TE13. Because each mode was a triplet that was not fully resolved, we fitted the data with a function that was the complex sum of 3 resonances plus a complex background, as described in Ref. [10]. The 3 peak frequencies for the resonances in the triplet, resulting from the fit, were averaged to determine the inner radius a of the BBB using the relation:

$$a = \frac{\xi_{1n}^{\sigma} c_0}{2\pi n_g \left\langle f_{1n}^{\sigma} \right\rangle} . \tag{7}$$

The refractive index $n_g(T, P)$ of the gas was obtained from Ref. [5]. The volume at 100 kPa and 295 K determined from the 3 microwave modes is

$$V_{\text{micro}} = 1.84743\left(1 \pm 1.1 \times 10^{-4}\right) \text{ m}^3 . \tag{8}$$

Equation (8) includes type A (random) and type B (systematic) uncertainties. The type B uncertainty is the standard deviation of the volume determined from 5 microwave modes (TM12, TM13, TE11, TE12, and TE13) and is an estimate of the effects of perturbations, such as non-sphericity and surface roughness, that are not included in our model of the microwave modes in the BBB. Figure 6 shows the agreement between V_{micro} and V_{gas} at 100 kPa and 295 K, $V_{\text{micro}}/V_{\text{gas}} - 1 = (2 \pm 14) \times 10^{-5}$.

We measured $V_{BBB}(P, T)$, the volume of the BBB as a function of temperature and pressure. The results are plotted in Fig. 1 and summarized by the fit in Eq. (1). The microwave measurements in nitrogen and in argon spanned the pressure range 100 kPa to 7 MPa and spanned laboratory temperatures between 293 K and 298 K. Equation (1) is valid over the rated working pressure range of the BBB. We expect, but cannot guarantee, the validity of Eq. (1) outside the measurement temperature range.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

4. Estimating the average temperature

4.1 Temperature sensor array

For a traditional *PVTt* primary gas flow standard, any errors in the gas temperature are directly transferred into errors in the calculated mass. The average temperature of the collected gas is difficult to measure, especially in un-thermostated environments. When pressurized gas flows into a large tank, the flow generates different temperatures in different parts of the tank. When the inflow stops, buoyancy-driven convection moves the warmest gas to the top of the tank and the coolest gas to the bottom. This stratification makes it difficult to measure the average temperature by conventional means. A prompt reading of a few thermometers is inherently inaccurate, and temperature gradients persist within the stagnant gas in big tanks.

As discussed in Section 3, during the gas expansion experiments, the BBB was instrumented with an internal thermocouple array consisting of ten thermocouples

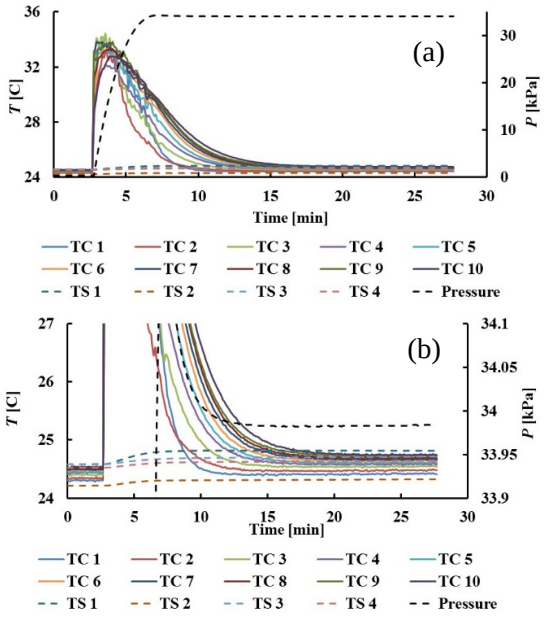


Figure 8. (a) Temperature and pressure profile inside the BBB. The solid lines are measurements from the internal thermocouple (TC) array and the dashed lines are the temperature measured by the surface thermistors (TS). (b) Expanded view of (a).

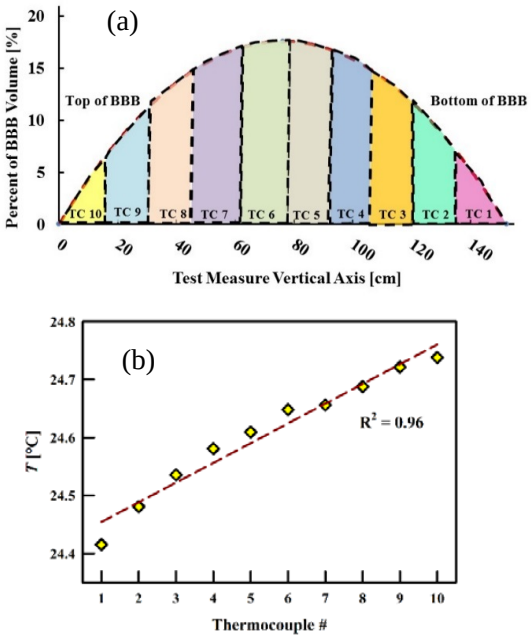


Figure 7. (a) Graphical representation of how thermocouple measurements are volume weighted. (b) Linear temperature distribution with respect to height in the BBB.

spaced at approximately 15.25 cm vertical intervals extending from the top of the BBB to its bottom. In addition, 4 surface thermistors were located at the top, the bottom, and on opposite sides of the equator of the BBB. We averaged the temperatures determined by the internal thermocouple array, weighting each thermocouple by the fractional volume of the BBB at its height. Figure 7(a) illustrates the volume weighting for a half sphere. Figure 7(b) shows the stratified temperature distribution for a quiescent state in the BBB while it was insulated. Usually, the temperature distribution in the BBB was a linear function of the height whether or not the BBB was insulated. When uninsulated, the quiescent temperature distribution spanned a larger temperature range (typically 1 K), reflecting the temperature distribution in the room.

The thermocouple array is advantageous because it captures the "fast" temperature transients in the BBB, whereas the surface thermistors do not. Figure 8 shows

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

the temperature profile in the BBB measured by the thermocouple array and the surface thermistors after pressurizing from < 1 kPa to approximately 30 kPa in less than 3 minutes.

The equilibration in Fig. 8 is roughly consistent with a simple model that ignores convective heat transfer. We consider the thermal equilibration of a spherical substance with radius a by radial heat conduction only [11]. The slowest time constant characterizing equilibration is $\tau = a^2/(D_T \pi^2)$, where D_T is the thermal diffusivity of the sphere. For argon at 295 K, $D_T \approx (0.7\ \mathrm{cm^2/s}) \times (P/30\ \mathrm{kPa})$ [5]. This model predicts $\tau \approx 14$ min for 30 kPa, which is approximately what we observed.

At 30 kPa, the surface thermistors did not capture the "fast" temperature transient generated by pressurizing the BBB; however, they did agree (within 80 mK) with the volume weighted average of the internal thermocouples after ≈20 minutes had passed. In these experiments, the BBB was insulated. However, the insulation was removed in subsequent experiments allowing the internal gas to more quickly equilibrate with the room. Remaining questions are: 1) how long does it take for the temperature to equilibrate following more extreme pressure changes? and 2) how accurate are the surface thermistors when the pressure and mass of gas in the BBB are two orders of magnitude larger than the conditions in Fig. 8 and the BBB is not insulated? For a *PVTt* primary standard to be practical, the thermal equilibration time and the accuracy of the temperature measurement must be acceptable. To answer these questions, we relied on the surface thermistors and measured acoustic resonance frequencies. We could not use the internal thermocouple array because the wires leading to the thermocouples interfered with the microwave measurements.

As discussed in Section 1 and shown in Fig. 2, the mass of gas deduced from the acoustic resonance frequencies and the internal gas pressure $M_{\mathrm{acoust}} \propto (P / f_{\mathrm{acoust}}^2)$ equilibrated significantly faster than the mass determined by the pressure and the average surface thermistor reading $P/(Z\langle T\rangle_{\mathrm{sh}})$. Furthermore, the masses determined by these methods long after a pressure change (at time infinity, M_∞) are inconsistent by as much as 0.25 % because the temperature in the laboratory is not controlled and changes in a diurnal cycle. (There is always at least a small discrepancy between the shell temperature and the internal gas temperature.) In the example in Fig. 2, M_{acoust} was determined within a 0.01 % uncertainty 8-times faster than the mass determined from the surface thermistors using $P/(Z\langle T\rangle_{\mathrm{sh}})$.

At this time, we have not completed reference measurements that validate M_{acoust}. In a future publication, we will report results utilizing a weighing rig to transfer 20 kg aliquots of argon into the BBB. Preliminary results indicate that $|M_{\mathrm{acoust}}/M_{\mathrm{grav}} - 1| < 5\times10^{-4}$.

4.2 Acoustic resonance frequency

Figure 5(a) shows the positions of the acoustic transducers in the BBB. The sound source, a voice-coil speaker with a hard-plastic cone, was mounted in the 5 cm-diameter port (Fig. 5(c)). A flexible foam material was placed behind the speaker to dampen the sound emitted from the rear of the speaker. The detector transducer (microphone) was mounted in the 2.5 cm-diameter port (Fig. 9). The detector was a 32 mm-diameter, 0.38 mm-thick, bimorph piezoceramic bending disk mounted to the flange with flexible caulk.

Previously, we have shown how the mass M of gas in a pressure vessel with volume V may be deduced from measurements of the gas pressure P and the frequencies f_{acoust} of acoustic resonances within the vessel. The mass

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

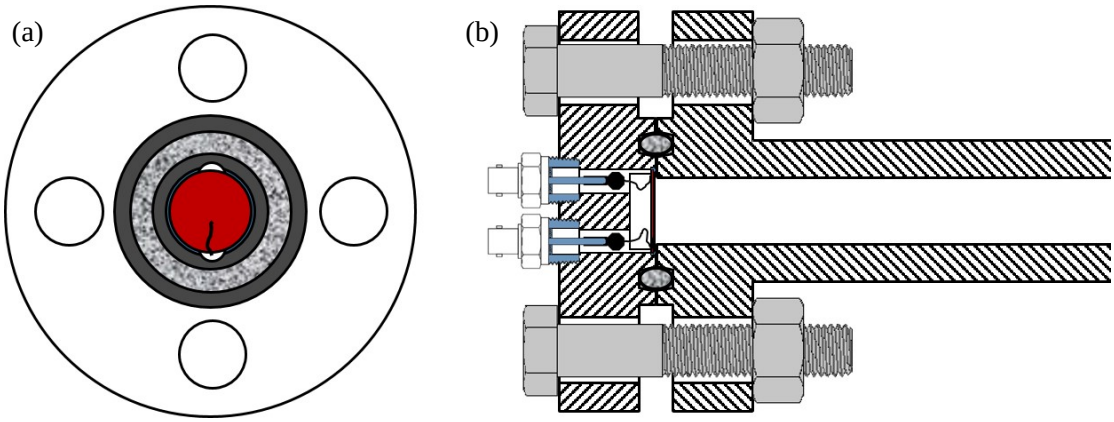


Figure 9. Schematic of the acoustic detector. (a) 32 mm diameter, 0.38 mm thick bimorph piezoceramic bending disk (red) mounted with flexible caulk in flange on the 2.5 cm diameter port. Small vents permit static pressure equilibration and connection to electrode. (b) Cross section view of assembled flanges and port duct. Two BNC, 3/8-NPT (10 MPa) feedthroughs are used for differential detection of the signal from the bimorph electrodes.

determined with this technique was shown to be insensitive to time-independent temperature gradients in the gas. The acoustic measurements eliminated the need for multiple thermometers to determine the average gas temperature with sufficient accuracy for flow metrology.

In a uniform gas, the speed of sound w is proportional to $\sqrt{T}$ plus small corrections, where T is the thermodynamic temperature of the gas. An acoustic mode of the gas in a pressure vessel is a standing wave with velocity potential $\varphi_i(\mathbf{r})$ and wavenumber k_i, where i denotes a set of indices that identify the mode; $\varphi_i(\mathbf{r})$ and k_i are gas independent. A driven sound wave in the confined gas has maximum amplitude at the resonance frequency $f_{\text{acoust}} \propto w$. The basis of acoustic gas thermometry is the inversion of the relationship to determine the gas's thermodynamic temperature from measured resonance frequencies.

When the temperature is not uniform, the speed of sound in the gas is spatially dependent. To first order, the resonance frequencies are determined by a mode-dependent, weighted average of $[w(\mathbf{r})]^2$ over the volume, where $w(\mathbf{r})$ is the local speed of sound. Thus, $f^2_{\text{acoust}} \propto \langle w^2 \rangle_\varphi$, where [7]

$$\langle w^2 \rangle_\varphi = \frac{\int_V w(\mathbf{r})^2 |\varphi_i|^2 \, dV}{\int_V |\varphi_i|^2 \, dV} \tag{9}$$

In spherical coordinates, the volume differential $dV = r^2 dr \, d\Omega$, where $d\Omega$ is a differential solid angle. The acoustic velocity potential φ_i is proportional to the local acoustic pressure in the standing wave. Regions where $\varphi_i = 0$, *i.e.* pressure nodes, do not contribute to the weighted average. (See Fig. 10.) In effect, $\langle w^2 \rangle_\varphi$ is proportional to the mode-dependent average temperature $\langle T \rangle_\varphi$. However, the average density of the gas in the vessel is dependent on the *volume* average of $1/T$. Because typical temperature variations δT are small, we can approximate $\langle 1/T \rangle$ by $1/\langle T \rangle$ with fractional errors of order $(\sigma / \langle T \rangle)^2 \approx 10^{-4}$, where σ is the standard deviation of the temperature. The remainder of

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

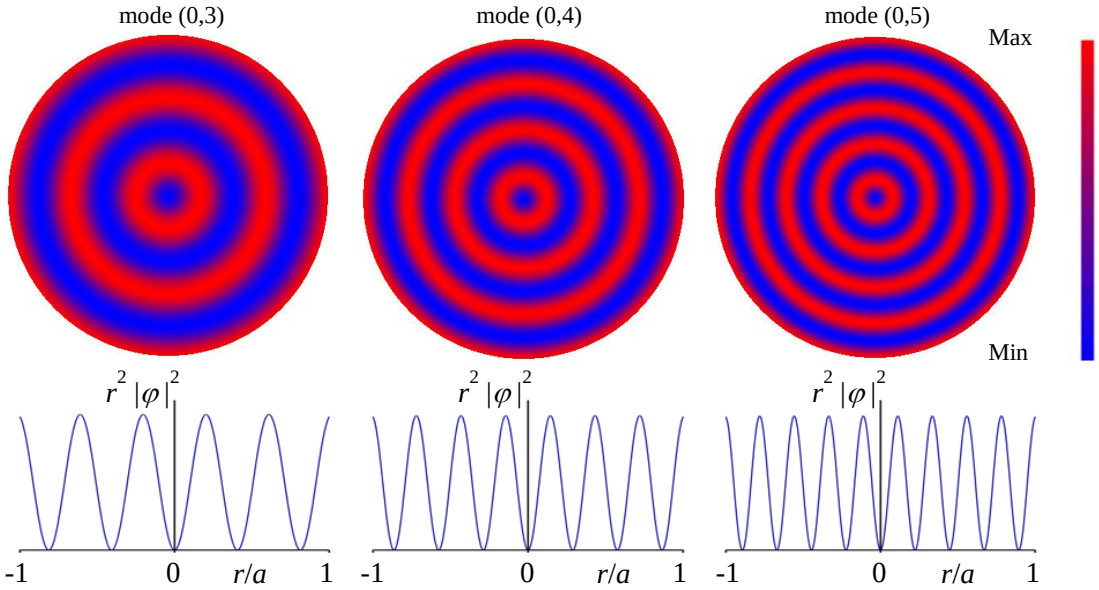


Figure 10. Top: Plots of the weighting function $r^2|\varphi|^2$ for 3 radial acoustic modes of a gas in a spherical cavity. Bottom: Graphs of $r^2|\varphi(r)|^2$ as a function of r/a. The plots are normalized by $\oint_V |\varphi|^2\, dV$ and have the same vertical scale.

this section explores the difference between $\langle T\rangle_\varphi$ and $\langle T\rangle_V \equiv (1/V)\int_V T(\mathbf{r})\, dV$ for simple temperature profiles in spherical and cylindrical pressure vessels and gives specific recommendations.

Gillis, *et al*. [6] showed that the measured resonance frequencies in a cylindrical vessel did not change to first order when a linear temperature gradient was imposed, while $\langle T\rangle_V$ was kept fixed, in agreement with first order perturbation theory, *i.e.* $\langle T\rangle_\varphi \equiv \langle T\rangle_V$ for a linear temperature gradient in a cylindrical vessel. The same is true in a spherical vessel. Figure 10 shows plots of $[r^2|\varphi(r)|^2]$ for three radial modes of a spherical cavity; the plots are normalized by the integral $\int_V |\varphi(\mathbf{r})|^2\, dV \equiv V\Lambda_\alpha$ and plotted on the same scale. (We use radial modes because they are non-degenerate and have the highest quality factors.) Because $[r^2|\varphi(r)|^2]$ is an even function about any plane through the sphere's center, a gradient in w^2 (or T) that is an odd function of the distance from that plane (*e.g.* linear) will integrate to zero in Eq. (9). Gradients that are not odd functions of the distance will, generally, have non-vanishing integrals.

During pressurization of the BBB, gas expanding past a throttle valve in the exterior plumbing cools by 20 °C or more before entering the BBB as a jet. Inside, the cold gas sinks to the lower region of the vessel, while the compressing gas heats up and rises to the top. The resulting density gradients and mass currents are too complex to model accurately. Instead, to compare the volume and mode averages, we considered the very simple step-function temperature distribution given by

$$T(z) = T_1 + \left[H\left(z - z_h\right) - 1\right]\Delta T\,, \qquad H\left(z - z_h\right) = \begin{cases} 0, & z < z_h \\ 1, & z > z_h \end{cases} \tag{10}$$

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

shown in Fig. 11. Here, $z_h \equiv -a + h$ is the z-coordinate for the location of the interface and $\Delta T = T_1 - T_2$. The volume-average temperature defined above becomes

$$\langle T \rangle_V = T_1 \frac{V_1}{V} + T_2 \frac{V_2}{V}, \tag{11}$$

which is a function of h. In a spherical geometry Eq. (11) becomes

$$\langle T \rangle_V = T_1 - \frac{3}{4}\frac{h^2}{a^2}\left(1 - \frac{1}{3}\frac{h}{a}\right)\Delta T . \tag{12}$$

The mode-average temperature obtained from Eq. (9) for the radial mode (0,n) becomes

$$\langle T \rangle_{\varphi_{0n}} = T_1 - \Delta T \int_{V_2} \frac{|\varphi_{0n}|^2}{V \Lambda_{0n}} dV \tag{13}$$

where the velocity potential is a function of r only

$$\varphi_{0n}(r) = j_0\left(\zeta_{0n}\, r/a\right), \text{ with } \Lambda_{0n} = \frac{3}{2}\left[\, j_0\left(\zeta_{0n}\right)\right]^2, \tag{14}$$

j_0 is the zeroth order spherical Bessel function, and the eigenvalue ζ_{0n} is the n^{th} root of $\zeta = \tan\zeta$. The averages $\langle\, T \,\rangle_V$ and $\langle\, T \,\rangle_\varphi$ from Eqs. (12) and (13) are plotted in Fig. 12(a) as a function of h/a for n = 2, 3, and 4. The fractional difference between these two averages is approximately

$$\frac{\langle T \rangle_{\varphi_{0n}} - \langle T \rangle_V}{\langle T \rangle_V} \approx \frac{\Delta T}{T_1}\left[\frac{3}{4}\frac{h^2}{a^2}\left(1 - \frac{1}{3}\frac{h}{a}\right) - \frac{2}{3V}\int_{V_2}\left(\frac{j_0\left(\zeta_{0n}\, r/a\right)}{j_0\left(\zeta_{0n}\right)}\right)^2 dV\right] \tag{15}$$

to first order in the small quantity $\Delta T / T_1$. For the extreme case considered in Figs. 2(a) and 2(b), ΔT = 20 K and T_1 = 300 K, $\Delta T / T_1 \approx 0.07$. For a more typical case (Fig. 2(c)) ΔT = 1.5 K and $\Delta T / T_1 \approx 0.005$ immediately after a gas collection.

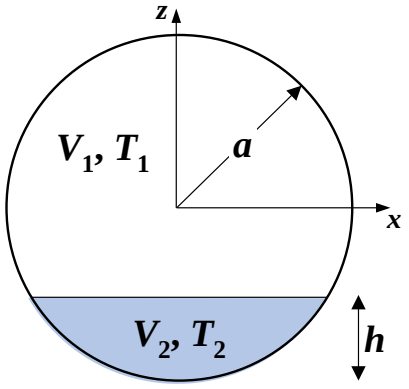


Figure 11. Simple temperature distribution in spherical vessel with radius a and volume $V = 4\pi a^3/3 = V_1 + V_2$. Upper region has volume V_1 and uniform temperature T_1; lower region with height h has volume V_2 and uniform temperature $T_2 < T_1$.

The quantity inside the square brackets in Eq. (15), which is a function only of h/a for a given radial mode n, is plotted in Fig. 12(b) for radial modes (0,n) with n = 2, 3, 4, and the limit of large n (dashed line). The central result from this simple model of temperature gradients in a spherical vessel is that the radial-mode-average $\langle w^2 \rangle_\varphi$, and therefore the density, may be in error, fractionally, by as much as $0.1 \times (\Delta T / T_1)$.

For completeness, we also investigated using non-radial modes to measure $\langle w^2 \rangle_\varphi$. Non-radial modes are identified by two integers (l,n) with $l > 0$ and have degeneracy $2l + 1$. The "simplest" non-radial mode is the (1,1) mode, and it is 3-fold degenerate. The three components of the

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

triplet have nodal (zero weight) planes that are mutually perpendicular and pass through the sphere's center. For one component the *xy*-plane is a nodal plane, and the greatest weight occurs in lobes above and below the nodal plane. Thus, this component will be most sensitive to gradients in the *z*-direction. The other two components have the *xz* and *yz*-planes as nodal planes. These components will do a better job of averaging over the gradient. Unfortunately, the acoustic microphone and source do not couple to the three components with equal efficiency, and there is no easy way to determine what average over these components is appropriate. For these and other reasons, we have chosen to measure the frequencies of only the (non-degenerate) radial modes.

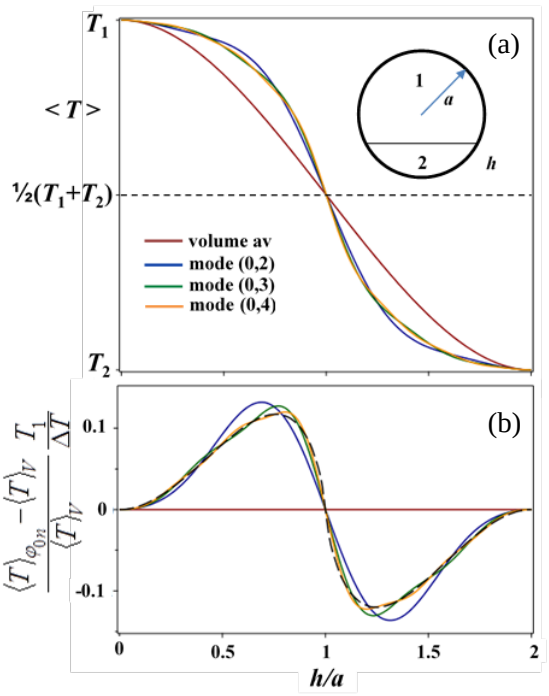


Figure 12. (a) The averages $\langle T \rangle_V$ and $\langle T \rangle_\varphi$ in a gas-filled spherical vessel from Eqs. (12) and (13) for n = 2, 3, and 4. (b) Scaled fractional difference between the volume and mode averages from Eq. (15). The black dashed curve is the large n limit.

We also investigated this model for a cylindrical geometry. For a cylindrical vessel, we need to consider the orientation of the cylinder's axis relative to the gradients that are likely to develop in earth's gravity, *i.e.* the gradient is a function of *z* only. For the reasons discussed above, we looked only at non-degenerate modes of the gas-filled cylindrical vessel, *i.e.* longitudinal and transverse radial modes, to determine $\langle w^2 \rangle_\varphi$. For a horizontal cylinder, the temperature gradient is perpendicular to the cylinder axis. $|\varphi(\mathbf{r})|^2$ for the longitudinal modes is only a function of the axial coordinate, so the temperature in V_1 and V_2 will be given equal weight. Therefore, the volume-average and longitudinal mode-average will be the same. Radial mode averages for a horizontal cylinder could be in error by as much as $0.1\times(\Delta T / T_1)$, similar to radial modes in spheres.

For vertically oriented cylindrical vessels, *i.e.* the gradient is parallel to the cylinder axis, the situation is reversed: $|\varphi(\mathbf{r})|^2$ for the radial modes is only a function of the coordinates perpendicular to the gradient, so the mode average is the same as the volume average. Longitudinal modes (parallel with the gradient) have averaging errors that diminish as $1/(2\pi l)$, where l is the longitudinal mode number. Table 3 summarizes the results of the linear and step temperature distributions.

Table 3. Fractional difference between $\langle T \rangle_V$ and $\langle T \rangle_\varphi$ for a linear gradient (dT/dz = constant) and the step function defined in Eq. (10) (Fig. 11). Results from first-order perturbation theory are given for a spherical resonator (radial modes) and a cylindrical resonator oriented horizontally or vertically using longitudinal and radial modes.

	Temperature Distribution	Spherical resonator	Cylindrical vessel	
			axis horizontal	axis vertical
linear	$T(z) = T_0 + (\mathrm{d}T/\mathrm{d}z)\, z$	0 (radial)	0 (longitudinal)	0 (radial)
step	$T(z) = T_1 + [H(z-z_h)-1]\,\Delta T$	$< 0.1\ (\Delta T/T_1)$ (radial)	$< 0.1\ (\Delta T/T_1)$ (radial)	0 (radial)
			0 (longitudinal)	$< (\Delta T/T_1)/2\pi l$ (long.)

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

5. Discussion

We characterized a 1.8 m^3, nearly-spherical, steel shell at pressures up to 7 MPa for use as a collection vessel or source for gas flow standard. We measured the cavity's microwave resonance frequencies f_{micro} to determine its pressure- and temperature-dependent volume: $V_{micro}(P, T) = 1.84740\ m^3 \times [1 + \alpha(T - 295\ K) + \kappa P]$ with a fractional uncertainty of 0.011 % at a 68 % confidence level. The coefficients α and κ were consistent with the dimensions and properties of the steel shell. The microwave-determined volume V_{micro} was consistent, within combined uncertainties, with V_{gas} the volume determined by a gas expansion method: $V_{micro}/V_{gas} - 1 = (2 \pm 14)\times 10^{-5}$.

When the shell was filled with gas, measurements of its acoustic resonance frequencies f_{acoust} and of the pressure quickly and accurately determined the mass of the gas in the shell, even when temperature gradients persisted. We have shown that despite the large thermal gradients that develop when gas is added or removed from the BBB, the mass M_{acoust} of gas determined from acoustic resonance frequencies settled to within 0.01 % of its final value much faster than the estimated mass determined from the shell temperature. Following a pressure change of 0.3 MPa, the top-to-bottom temperature difference was 1.5 °C and M_{acoust} settled to its final value in just 0.5 h. Proportionately longer times were required after larger pressure changes generated larger temperature differences. In a subsequent publication, we will compare the mass M_{acoust} of gas in the BBB determined from acoustic measurements with M_{grav} determined gravimetrically. This comparison will determine the accuracy with which we can measure the mass of gas from acoustic measurements. Our preliminary measurements indicate $|M_{acoust}/M_{grav} - 1| < 5\times 10^{-4}$.

Finally, based on our study of acoustic measurements in collection vessels with vertical temperature gradients, cylindrical vessels may have advantages over spherical vessels. Measurements of $\langle w^2 \rangle$ and, therefore, $\langle \rho \rangle$ from the longitudinal acoustic modes of a gas-filled, horizontal cylindrical vessel should be insensitive to vertical temperature gradients, contributing an uncertainty to the mass of order $(\Delta T/T)^2$. For a vertical cylinder, the uncertainty due to vertical temperature gradients using longitudinal modes are first order in $(\Delta T/T)$, but decrease as the mode order increases. The uncertainty for radial acoustic modes in a vertical cylindrical vessel should be second order.

6. References

[1] A.N. Johnson. "Natural Gas Flow Calibration Service" NIST special publication 1081, August 2008.

[2] J.D. Wright, A.N. Johnson, M.R. Moldover, and G.M. Kline "Gas Flowmeter Calibrations with the 34 L and 677 L *PVTt* Standards" NIST Special Publication 250-63, November 2010.

[3] A.N. Johnson, C. Li, J.D. Wright, G.M. Kline, and C.J. Crowley. "Improved Nozzle Manifold for Gas Flow Calibrations", 8th International Symposium on Fluid Flow Measurement. Colorado Springs, CO, U.S.; 6/18/2012 to 6/22/2012.

[4] J.P.M. Trusler, *Physical Acoustics and Metrology of Fluids* (Adam Hilger, Bristol, 1991).

[5] E.W. Lemmon, M.O. McLinden, and M.L. Huber, "REFPROP: Reference Fluid Thermodynamic and Transport Properties," NIST Standard Reference Database 23, Version 9.1, Natl. Inst. Stand. and Tech., Boulder, CO, (2010) www.nist.gov/srd/nist23.cfm

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

[6] K.A. Gillis, J.B. Mehl, J.W. Schmidt, and M.R. Moldover. "'Weighing' a gas with microwave and acoustic resonances", Metrologia. **52**, 337-352 (2015).

[7] K.A. Gillis, M.R. Moldover, and J.B. Mehl, "Detecting Leaks in Gas-Filled Pressure Vessels Using Acoustic Resonances", Rev. Scientific Instrum. **87**, 054901 (2016).

[8] M.R. Moldover, J.W. Schmidt, K.A. Gillis, J.B. Mehl, and J.D. Wright. "Microwave determination of the volume of a pressure vessel", Meas. Sci. Technol. **26**, 015304 (2015).

[9] J.B. Mehl and M.R. Moldover, "Measurement of the ratio of the speed of sound to the speed of light," Phys. Rev. A **34**, 3341-3344 (1986).

[10] E.F. May, L. Pitre, J.B. Mehl, M.R. Moldover, and J.W. Schmidt, "Quasi-spherical cavity resonators for metrology based on the relative dielectric permittivity of gases," Rev. Sci. Instrum. **75**, 3307-3317 (2004).

[11] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids: 2nd Edition* (Clarendon Press, Oxford, Great Britain, 1959), Section 9.3.

Gillis, Keith; Mehl, J; Moldover, Michael; Pope, Jodie.
"Characterizing Gas-Collection Volumes with Acoustic and Microwave Resonances."
Paper presented at 10th annual International symposium for fluid flow measurements, ISFFM, Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing

Iosif Shinder[1], Aaron Johnson, Michael Moldover, James Filla, Vladimir Khromchenko

Sensor Science Division, NIST, Gaithersburg, MD 20899

NOMENCLATURE

Variables with Roman Letters

$a_{q,0}$ - zeroth degree polynomial fit coefficient of null data (equals Γ^{q}_{NULL})

c_0, c_1, c_2 - fit coefficients for density correlation

C_{LDA} - calibration coefficient for NIST's Laser Doppler Anemometry (LDA) system (Usually C_{LDA} = 1.)

$F_1 = \Delta P_{45,NULL} / \Delta P_{12,NULL}$ - pitch calibration factor

$F_2 = \sqrt{\rho V_{LDA}^2 / 2\Delta P_{12,NULL}}$ - velocity calibration factor

$\vec{n}_{P1}$ - unit vector normal to port 1 on five-hole probe

T - static temperature of the air in the wind tunnel

P - static pressure of the air in the wind tunnel

P_{offset} - yaw pressure at which null parameters are measured (may be offset from zero)

RH - relative humidity of the air in wind tunnel

$PR_1 = \Delta P_{45} / \Delta P_{12}$ - Pitch pressure ratio

$PR_2 = \sqrt{\Delta P_{dyn} / \Delta P_{12}}$ - Velocity pressure ratio

$r(x_m, x_n)$ - normalized correlation coefficient between variables x_m and x_n

$u(x)$ - uncertainty of input quantity x at the 68 % confidence level

$u_{c,p}(y)$ - combined propagated uncertainty of measurand y at the 68 % confidence level

$u_{c,np}(y)$ - combined non-propagated uncertainty of measurand y at the 68 % confidence level

$u_c(y)$ - combined uncertainty of measurand y at the 68 % confidence level

$U(y)$ - expanded uncertainty of measurand y at the 95 % confidence level

V_{LDA} - airspeed measured by NIST's Laser Doppler Anemometry (LDA) working standard

V_{PITOT} - airspeed measured by NIST's standard L-type pitot probe check standard

$V_{z'}$ - local velocity along stack axis

Variables with Greek letters

α - pitch angle measured by traverse system

Γ_q - global probe variable equals a null parameter or calibration coefficient depending on value of q

$\Delta P_{dyn} = \rho V_{LDA}^2 / 2$ - dynamic pressure based on air density and airspeed

1 Corresponding Author: Iosif.shinder@nist.gov

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$\Delta P_{mn} = P_m - P_n$ - measured differential pressure across ports "m" and "n" of the five-hole

- $\Delta P_{45} = P_4 - P_5$ - *pitch pressure*
- $\Delta P_{23} = P_2 - P_3$ - *yaw pressure*
- $\Delta P_{12} = P_1 - P_2$ - *velocity pressure*

ΔP_{zero} - zero reading of differential pressure transducer

ΔP_{cal} - differential pressure reading within calibrated range transducer

ε - positive number much less than 1

ρ - air density

θ - yaw angle measured by traverse system

Ψ - rational polynomial curve fit of calibration data

Subscripts

ALIGN - alignment of yaw angle or pitch angle during probe installation

EOS - Equation of State

FIT - curve fitted value

LDA - Laser Doppler Anemometry

NULL- condition where yaw pressure is zero ($\Delta P_{23} = 0$)

q - global probe parameter with values 1, 2, 3, 4, or 5

r - relative value

TRAV - wind tunnel traversing system

x – axis of wind tunnel

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

ABSTRACT

We report progress towards the goal of reducing the errors in industrial smokestack flow measurements to 1 % by replacing S-probes with calibrated 3-D probes (i.e., probes that measure 3 components of velocity). NIST calibrated a commercially-manufactured spherical probe and a prism probe at air speeds (5 m/s to 30 m/s) and pitch angles (–20° to 20°) using a yaw-nulling method similar to EPA's Method 2F. The expanded uncertainty for 3-D air speed measurements of both probes was near 1 % at a 95 % confidence level. Most of this uncertainty is attributed to the reproducibility of the calibration measurements and the uncertainty of the NIST's laser doppler anemometer air-speed standard. Thus, the 1 % goal might be possible; however, to obtain this low uncertainty, two significant uncertainties consistent with EPA Method 2F must be avoided. First, the Reynolds number dependence of the prism probe must be accounted for during probe calibration, and second, more stringent uncertainty requirements are needed for the yaw-pressure measurement at low flows.

1 INTRODUCTION

Accurate stack-flow measurements are required to measure hazardous emissions from industrial smokestacks. Today, most stack flow measurements use an S-type pitot probe (or S-probe), which measure only two components of the flow velocity. When the velocity field in stacks is highly 3-dimensional, S-probe surveys can have errors of 10 % or more [1 – 4]. Here, we report progress towards the goal of reducing stack flow survey errors to 1 % by replacing S-probes with calibrated probes that measure 3 components of velocity.

The U.S. Environmental Protection Agency (EPA) has developed regulations supporting two types of commercially available 3-D probes including a 5-hole prism probe and a 5-hole spherical probe shown in Figures 1A and 1B. These probes are calibrated and used in accordance with EPA document Method 2F [5]. The protocols in Method 2F stipulate 1) the design features of the wind tunnel used to calibrate probes (*e.g.*, minimum cross section dimensions, flow stability, range of air speeds, maximum level of off-axis flow), 2) the calibration procedure to determine the Method 2F calibration factors, aptly named F_1 and F_2, 3) the procedure for using the probe to measure the volumetric flow in industrial stacks, and 4) the uncertainty specifications of auxiliary measurements of differential pressure, yaw angle, and pitch angle. However, the uncertainty of the calibration factors F_1 and F_2 are not specified. Therefore, it is not clear what level of uncertainty reduction, if any, will be gained by replacing an inexpensive two-dimensional (2-D) S-probe with a more complex three-dimensional (3-D) probe. If 1 % stack flow measurements are the goal, then the uncertainty of these calibration factors should be no larger than 1 % at the 95 % confidence level.

The National Institute of Standards and Technology (NIST) has experience calibrating both prism probes and spherical probes using nulling and non-nulling methods. For each probe type, the calibration factors have the same general characteristics. In this work we calibrated the prism probe and the spherical probe shown in Figures 1A and 1B using the Method 2F protocol; however, NIST implementation of Method 2F exceeded the protocol's minimum requirements. Several steps were taken to obtain the highest quality data with the lowest possible uncertainty. First, we characterized the probes' Reynolds number (*Re*) dependence over the full velocity range (5 m/s to 30 m/s) of typical coal-fired power plant stacks. In addition, we evaluated the probe

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

performance over a wide range of pitch angles (α = - 45° to 45°). The calibration was performed in NIST's wind tunnel [1] against the U.S. national standard for wind speed, a laser doppler anemometer (LDA) traceable to the SI unit of velocity *via* length and time [6].

Second, we computed the expanded uncertainties[2] of the both the pitch calibration factor F_1 and the velocity calibration factor F_2 for both 3-D probes. The results showed the expanded uncertainty (95 % confidence level) of the velocity calibration factor $100 \times U(F_2)/F_2$ ranged from 0.9 % to 1.1 % for the spherical probe and from 0.9 % to 1.25 % for the prism probe for pitch angles α = - 20° to 20°. We performed multiple calibrations on each probe and included reproducibility in the uncertainty budget. For both probes reproducibility was a major factor in the uncertainty budget, contributing more than 50 % to the total uncertainty. Given that the uncertainties of these probes have not been documented over the wide range of pitch angles and velocities investigated herein, these results will give the stack flow measurement community the groundwork for computing the uncertainty as well as provide a baseline of the best possible performance once can expect from these devices.

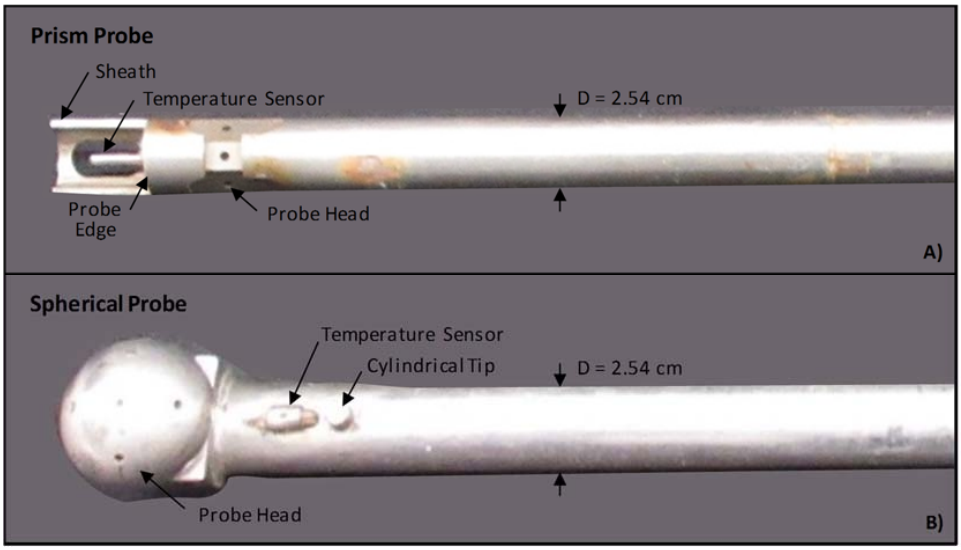


Figure 1. Picture of A) 5-hole prism probe and B) 5-hole spherical probe. The nominal diameter and probe length of each probe was D = 2.54 cm (1 inch) and 1.83 m (6 ft).

Third, we developed robust correlations that calculate the probe velocity calibration factor (F_2) and pitch angle (α) as a function of the measured pitch pressure and velocity pressure at the yaw-null angle. The correlation accounts for the probes' *Re* − dependence and pitch angle response over the range of pitch angles, $F_2 = F_2(\alpha, Re)$. For α = - 20° to 20° we found that $F_{2,PRISM}$ had a strong *Re*−dependence while $F_{2,SPHERICAL}$ was nearly independent of *Re*. Because the Method 2F protocol allows probes to be calibrated at only two air speeds, 18.3 m/s and 27.4 m/s, the resulting F_1 and F_2 calibration factors are extrapolated to lower velocities during application. This extrapolation procedure works reasonable well for the spherical probe since $F_{2,SPHERICAL}$ is nearly independent of

2 Herein the expanded uncertainty is the uncertainty with a confidence interval of 95 % or equivalently with a coverage factor of two (k = 2). A capital U is used throughout to denote expanded uncertainty.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Re. A reference value of $F_{2,SPHERICAL}$ measured at 20 m/s deviated less than ± 1 % with $F_{2,SPHERICAL}$ at other air speeds; the only exception occurred at the lowest air speed of 5 m/s where the deviation increased to nearly 2 %. However, for the prism probe the extrapolation introduces a 6 % error at low air speeds (< 10 m/s). To accurately measure low velocities a prism probe must be calibrated at low velocities so that the *Re*−dependence of $F_{2,PRISM}$ is adequately characterized.

Finally, we ensured that the accuracy of low speed (< 10 m/s) yaw pressure measurements only made a negligible contribution to the F_1 and F_2 uncertainty budget. When Method 2F protocol is implemented the uncertainty of the yaw pressure measurement can be significant at low speeds for the following 3 reasons. First, Method 2F only requires that the yaw pressure measurement be 1 % of full-scale where the full-scale is 124.4 Pa. As such, the accuracy of the yaw pressure measurement can be a significant fraction of the of the dynamic pressure at low air speeds. Second, during a flow RATA, Method 2F allows the differential pressure transducer that measures yaw pressure to drift by as much as 7.5 Pa. At low flue gas velocities, the uncertainty attributed to the zero drift can be a significant fraction of the dynamic pressure, leading to a large uncertainty in the yaw pressure measurement. Finally, Method 2F requires *nulling the probe* or equivalently finding the yaw angle resulting in zero yaw pressure at each traverse point. In practice, finding the probe yaw-angle where the yaw-pressure is zero is difficult due to noisy pressure signals. This is especially true in typical stack applications that have swirling, turbulent flows. Method 2F does not require time-averaging the yaw pressure measurements and adjusting the yaw angle until the actual zero yaw pressure is determined. Consequently, the F_1 and F_2 values determined using Method 2F are generally measured at a yaw-pressure close to, but not exactly equal to zero. This uncertainty has not been documented in Method 2F, but it can be significant at low velocities.

To mitigate the large yaw pressure uncertainties that occur at low velocities we took the following 3 steps. First, we measured the yaw-pressure with a stable, low uncertainty yaw-pressure transducers with uncertainty specifications based on 0.2 percent of reading instead of 1 % percent of full-scale. Second, we ensured that zero drift was negligible relative to other uncertainty components during calibration. Finally, we implemented a curve fit method (CFM) to ensure that the calibration factors F_1 and F_2 are determined at a zero yaw-pressure, even when pressure signals are noisy due to flow noise.

This manuscript documents the calibration procedures used by NIST, gives the results for both 3-D probes, and documents their uncertainty calculations. We hope that this paper will guide the stack flow community in estimating the uncertainty of Method 2F calibrations. The low uncertainties obtained herein are generally not obtainable using Method 2F unless certain modifications are made to the protocol. We conclude the manuscript by discussing modifications that should be made to Method 2F to avoid unexpected large uncertainties.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

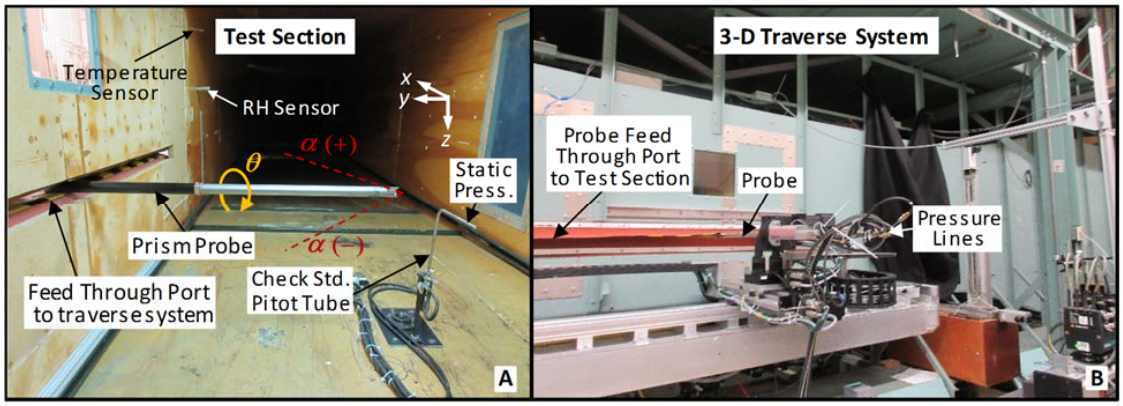


Figure 2. 3-D probe installed in NIST wind tunnel: A) definitions of the pitch (α) and yaw angles (θ), B) NIST 3-D Automated Traverse System. (Note that the air flow moves in the direction of the x-axis in Fig. 2).

2 DESCRIPTION OF NIST'S WIND TUNNEL

Air speed measurements were performed in the 2 m long rectangular test section of our recirculating wind tunnel [7 - 10]. The cross section is 1.2 m high by 1.5 m wide. This large cross-section minimizes the probe-induced blockage effects and the profile effects attributed to the boundary layer along the wall. LDA surveys (without a probe installed) demonstrate a uniform velocity in the test section's measurement zone. Figure 2A shows a 5-hole prism probe installed in the test section. Since the presence of the probe disturbs the velocity field, the LDA velocity (V_{LDA}) is measured in the freestream upstream of the probe's zone of influence.

The velocity measured by the LDA is converted to dynamic pressure by

$$\Delta P_{dyn} = \frac{\rho V_{LDA}^2}{2} \tag{1}$$

where ρ is the air density in the test section, which is calculated by

$$\rho = \frac{c_0}{T}\left[P - c_1 RH \exp\left(-\frac{c_2}{T}\right)\right] \tag{2}$$

where the coefficients have the following values $c_0 = 3.4848\times10^{-3}$ K kg/J, $c_1 = 6.65287\times10^{8}$ Pa, and $c_2 = 5315.56$ K; RH is the relative humidity of the air in percent, P is the static pressure in Pascal, T is air temperature in Kelvin, and the air density is in units of kg/m^3. The pressure is measured using the static pressure ports on the NIST check standard pitot probe shown in Fig.2A. The figure also shows the location where the temperature and relative humidity are measured.[3] The standard relative uncertainty of density equation of state is $u_{EOS,r}(\rho) = 0.1$ % [11].[4] The standard uncertainties of the temperature, pressure, and relative humidity (RH) are $u(T) = 1$ K, $u(P) = 0.1$ kPa, and $u(RH) = 5$ %, respectively. These uncertainties are multiplied by their

3 In Fig. 2 the temperature sensors attached to the probes were not the ones used in this calibration.

4 Throughout this manuscript we denote relative uncertainties using the subscript "r", so that $u_r(x) = 100\, u(x)/x$ is a dimensionless value expressed as a percent; here, x is the measurand and $u(x)$ is the dimensional standard uncertainty of the measurand at a confidence interval of 68 % or equivalently with a coverage factor of unity ($k = 1$).

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

respective normalized sensitivity coefficients and root-sum-squared to give the relative uncertainty in the computed density.[5] For the range of air speeds in this calibration (5 m/s to 30 m/s) the expanded uncertainties of the LDA and L-shaped pitot probe are $U_{r,LDA}$ =0.41 % and $U_{r,L\text{-}PITOT}$ = 0.44 % [6], respectively.[2] During the calibration the average difference between the LDA airspeed (V_{LDA}) and the L-shaped pitot probe air speed (V_{PITOT}) differed by no more than 0.1 %, well within their expected uncertainties.

Probes are oriented to specific pitch angles in the range $-45° \leq \alpha \leq 45°$ and yaw angles in the range $-180° \leq \theta \leq$ to $+180°$ by the 3-D traverse system attached to the outside of the wind tunnel shown in Fig. 2B. The probe is installed into the test section through the narrow rectangular slot called the feed-through port. Velocity surveys have shown that air leakage into the test section through the feed through has a negligible effect on air speed measurements, provided the probe is positioned near the center of the wind tunnel. As shown in Fig. 2A the yaw angle (θ) corresponds to rotations about the probe axis while the pitch angle (α) rotates the probe in the x-y plane about the pressure-sensing port 1 (see Fig. 3 for pressure port locations). Consequently, when the traverse system moves the probe to a specified pitch angle the physical location of port 1 in the test section remains unchanged. In addition, the physical location of the LDA velocity measurement coincides with the streamline normal to port 1. The positive yaw angle direction is indicated by the direction of the arrow in Fig. 2A and the positive and negative pitch angles are denoted by the respective positive sign (+) and negative sign ($-$).

A custom software program controls the air speed and the probe pitch and yaw angle settings. The program also acquires the calibration data including the LDA velocity, the required differential pressure measurements, pitch, yaw, etc. Data acquisition uses several different communication interfaces. Pressure-based instrument readings are acquired using a PCI-based multifunction DAQ board, while auxiliary LabVIEW programs are used to continuously monitor the Laser Doppler Anemometer and environmental sensors (*i.e.*, temperature, pressure, and relative-humidity). The program ensures flow is stable and within ±0.2 % of set point velocity during data collection.

5 Section 5.1 gives details for calculating sensitivity coefficients and performing uncertainty analysis.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

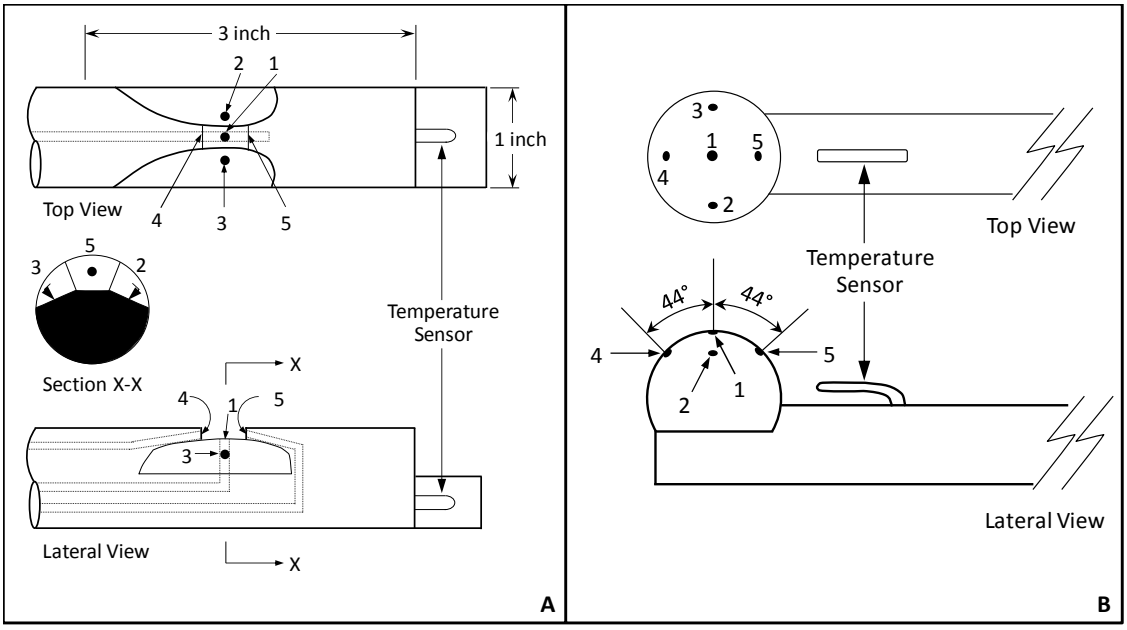


Figure 3. Probe head geometries and port locations 1, 2, 3, 4, and 5 for Prism probe (A) and Spherical probe (B).

3 PROBE DESCRIPTION AND INSTALLATION

3.1 Differential Pressure Measurements

Figures 3A and 3B show the shapes of the prism probe head, the spherical probe head, and the location of the 5 pressure ports on each probe. The pressure ports are labeled 1, 2, 3, 4, and 5. We measured the differential pressures across the ports using four 10 Torr (10 Torr = 1333.22 Pa) model 698A Baratron heated, high accuracy, bidirectional differential capacitance manometers.[6] For convenience, we define the pressure difference between any two ports by

$$\Delta P_{\mathrm{mn}} = P_{\mathrm{m}} - P_{\mathrm{n}} \tag{3}$$

where P_{m} is the pressure of port "m" where m = 1, 2, 3, 4, or 5 and P_{n} is the pressure of port "n" where n =1, 2, 3, 4, or 5. We connected the four pressure sensors to the pitot probe ports so that the first sensor measured ΔP_{12}; the second transducer measured ΔP_{13}; etc. The standard uncertainty of each differential pressure transducer is $u_{\mathrm{r}}(\Delta P_{\mathrm{cal}})$ = 0.1 % of reading for ΔP_{cal} = 2 Pa to 666.7 Pa and from $\Delta P_{\mathrm{cal}} = -2$ Pa to −666.7 Pa.[7] A conservative estimate of the standard uncertainty attributed to zero drift is conservatively taken to be $u(\Delta P_{\mathrm{zero}})$ = 0.002 Pa. Thus, the uncertainty of differential pressure measurements is

6 Certain commercial equipment, instruments, or materials are identified in this report to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

7 The uncertainty is based on repeated calibrations against NIST's primary pressure standard. The stated uncertainty includes the standard deviation of fit residuals, hysteresis, reproducibility, and the uncertainty from the primary pressure standard.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$u(\Delta P_{1n}) = \sqrt{u^2(\Delta P_{cal}) + u^2(\Delta P_{zero})} \quad (4A)$$

for n = 2 to 4, where ΔP_{1n} is the measured differential pressure, and $u(\Delta P_{cal})$ is the dimensional uncertainty (in units of pressure) attributed to the calibration of the transducer.

The pressure difference between ports 1 and 2 (ΔP_{12}) has special significance and is denoted "*velocity pressure*". Two other important differential pressure measurements include the "*yaw pressure"* and the "*pitch pressure"*. The yaw pressure is the differential pressure between ports 2 and 3 (ΔP_{23}) while the pitch pressure is the differential pressure between ports 4 and 5 (ΔP_{45}). Since NIST measured all the probe differential pressures relative to port 1, we subtract the appropriate differential pressure measurements to determine the yaw pressure and the pitch pressure. The yaw pressure is ΔP_{13} - ΔP_{12} and the pitch pressure is ΔP_{15} - ΔP_{14}.

The uncertainty in the velocity pressure measurement (ΔP_{12}) is given by Eq. (4A). However, additional considerations are necessary to determine the uncertainty of the yaw pressure (ΔP_{23}) and the pitch pressure (ΔP_{45}) since each is determined by the difference of two separate measurements. The yaw pressure will be zero at the yaw-null condition so that the two pressure measurements ΔP_{13} and ΔP_{12} will be nearly identical. In this case, the uncertainty from calibrating both transducers to the same standard is perfectly correlated, and does not contribute to the uncertainty of ΔP_{13} and ΔP_{12}. In contrast, the uncertainty from zero drift is not correlated and must be included for both transducers. Consequently, the uncertainty for the yaw pressure at (or near) the yaw condition is

$$u(\Delta P_{23}) = \sqrt{2}\, u(\Delta P_{zero}) \quad (4B)$$

where the factor $\sqrt{2}$ accounts for the zero drift of both transducers. For the uncertainty in the pitch pressure, the two measurements ΔP_{15} and ΔP_{14} are not equal, and their uncertainties of their calibrations are not correlated. In this case the uncertainty is

$$u(\Delta P_{45}) = \sqrt{2u^2(\Delta P_{cal}) + 2u^2(\Delta P_{zero})} \quad (4C)$$

where $\sqrt{2}$ is a factor of both $u(\Delta P_{cal})$ and $u(\Delta P_{zero})$.

We leak checked each pressure line before calibration to prevent erroneous pressure readings resulting from leaks.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

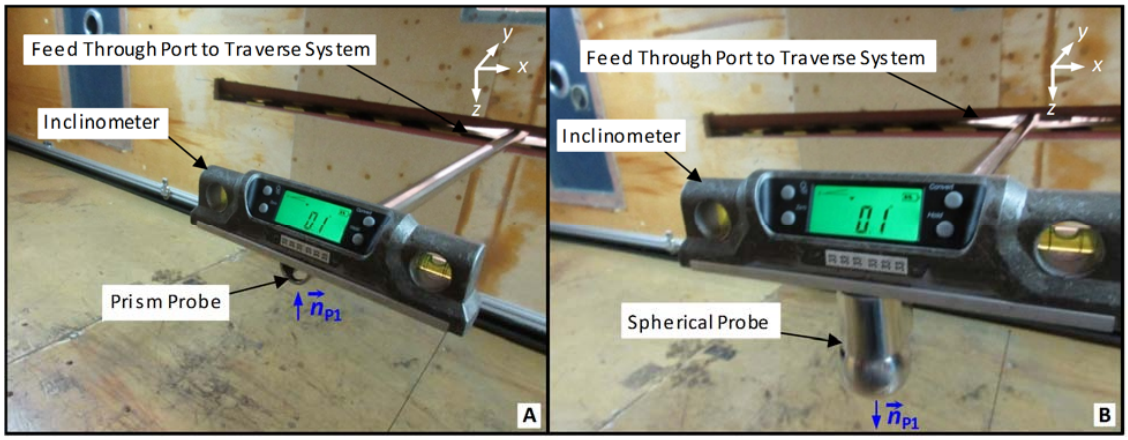


Figure 4. Yaw angle alignment: (A) Prism probe, and (B) Spherical probe in wind tunnel test section. Flow moves in the positive x-direction.

3.2 Alignment of 5-hole Probes

Figures 4A and 4B show pictures of the yaw angle alignment of both the prism probe and the spherical probe in the wind tunnel test section. The inclinometer in each figure is balanced on the flat sections of each probe head. The probe shaft is rotated until the inclinometer reads zero within ± 0.5°. At these reference yaw angles, the vector normal to port 1 (shown as $\vec{n}_{P1}$ in the figures) is orthogonal to the axis of the wind tunnel (the x-axis). In Fig. 4A port 1 on the prism probe is oriented upwards in the direction of $\vec{n}_{P1}$ so that the yaw angle is –90°. In contrast, port 1 on the spherical probe shown in Fig. 4B is oriented downward in the direction of $\vec{n}_{P1}$ so that the yaw angle is 90°. Based on this convention the yaw angle is zero when the probe is rotated so that $\vec{n}_{P1}$ points into the oncoming flow (*i.e.*, $\vec{n}_{P1}$ is parallel to but opposite the x-axis).

We define the pitch angle equal to zero when the probe axis is parallel to the transverse direction depicted by the y-axis in Fig. 4A and 4B. The traverse system is used to align probe to the zero-pitch position. The standard uncertainty of rotating a probe to the zero-pitch position and to all pitch angles ± 45° is $u(\alpha)$ = 0.25 °.

The standard uncertainty of the yaw angle is $u(\theta)$ = 0.32°. The yaw angle uncertainty is given by

$$u(\theta) = \sqrt{u^2_{\text{ALIGN}}(\theta) + u^2_{\text{TRAV}}(\theta)} \tag{5}$$

where $u_{\text{ALIGN}}(\theta)$ is the standard uncertainty attributed to yaw angle misalignment incurred during probe installation, and $u_{\text{TRAV}}(\theta)$ is the standard uncertainty associated with the traversing system's rotary encoder. The standard uncertainty attributed to yaw angle misalignment is $u_{\text{ALIGN}}(\theta)$ = ± 0.5°/$\sqrt{6}$ ≅ ± 0.2° where ± 0.5° is accuracy of the inclinometer shown in Fig. 4. Assuming a triangular distribution, the standard uncertainty is obtained by dividing ± 0.5° by $\sqrt{6}$ [12].

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

4 CALIBRATION PROCEDURES AND RESULTS

We calibrated both the spherical and prism probes using the framework of the Method 2F procedure. Deviations from the Method 2F procedure are discussed in detail. The spherical probe was calibrated on 3 different occasions: 7/12/2016, 7/15/2016, and 11/1/2016 while the prism probe was calibrated on 2 occasions: 7/18/2016 and 7/22/2016. The calibration conditions (*i.e.*, the range of air speeds, pitch angles, and yaw angles) are listed in Table 1 for the spherical probe and in Table 2 for the prism probe.

Table 1*.Calibration conditions for the 5-hole spherical probe.*

No.	Date	Range of LDA Velocities (V_{LDA})	Range of Pitch Angles (α)	Range of Yaw Angles (θ)
1	7/12/2016	5 m/s to 30 m/s (5 m/s steps)	–45° to 45° (3° steps)	–9° to 9° (3 ° steps, N = 7)
2	7/15/2016	7.5 m/s to 27.5 m/s (5 m/s steps)	–43.5° to 43.5° (3 ° steps)	–7.5° to 7.5° (3 ° steps, N = 6)
3	11/1/2016	5 m/s to 30 m/s (5 m/s steps)	–45° to 45° (5 ° steps)	–10° to 10° (2 ° steps, N = 11)

Table 2. *Calibration conditions for the 5-hole prism probe.*

No.	Date	Range of LDA Velocities (V_{LDA})	Range of Pitch Angles (α)	Range of Yaw Angles (θ)
1	7/18/2016	5 m/s to 30 m/s (5 m/s steps)	–45° to 45° (3 ° steps)	–9° to 9° (3 ° steps, N = 7)
2	7/22/2016	7.5 m/s to 27.5 m/s (5 m/s steps)	–43.5° to 43.5° (3 ° steps)	–7.5° to 7.5° (3 ° steps, N = 6)

After installing the probe into the traversing system and leak checking each pressure line, the air speed in the wind tunnel is set to the lowest-velocity set point and allowed to stabilize. While the air speed is stabilizing, the traverse system moves the probe to the initial pitch and yaw angles. After the pressure readings and the air speed are stable, we record the differential pressure readings from the probe (*i.e.*, ΔP_{12}, ΔP_{13}, ΔP_{14}, and ΔP_{15}), the LDA velocity (V_{LDA}), the static pressure, static temperature, and relative humidity. Subsequently, the probe is rotated to the next yaw angle and another set of data is recorded. We repeat this procedure until data is acquired at all the prescribed yaw angles. After traversing the range of pitch angles, we repeat the entire procedure from the beginning at the next air speed set point.

4.1 Null Parameters and Calibration Factors

At each air speed and pitch angle, we measured the yaw-null angle (θ_{NULL}) by determining the yaw angle for which the yaw pressure is zero (*i.e.*, $\Delta P_{23} = 0$). At the yaw-null angle we also measured the null pitch pressure ($\Delta P_{45,NULL}$) and the null velocity pressure ($\Delta P_{12,NULL}$). These null-parameters are used in conjunction with the measured air speed (V_{LDA}) and air density (ρ) to

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

determine the *pitch angle calibration factor* (F_1) and the *velocity calibration factor* (F_2). The formulas used to compute the pitch angle calibration factor and the velocity calibration factor are:

$$F_1 = \frac{\left(P_4 - P_5\right)_{\text{NULL}}}{\left(P_1 - P_2\right)_{\text{NULL}}} = \frac{\Delta P_{45,\text{NULL}}}{\Delta P_{12,\text{NULL}}}, \tag{6A}$$

and

$$F_2 = C_{\text{LDA}} \sqrt{\frac{\Delta P_{\text{dyn}}}{\left(P_1 - P_2\right)_{\text{NULL}}}} = C_{\text{LDA}} \sqrt{\frac{\Delta P_{\text{dyn}}}{\Delta P_{12,\text{NULL}}}} = \sqrt{\frac{\rho V_{\text{LDA}}^2}{2\Delta P_{12,\text{NULL}}}}, \tag{6B}$$

respectively. In Eq. (6B) the dynamic pressure is calculated from Eq. (1) so that the coefficient C_{LDA} is identically equal to 1.

4.2 Curve Fit Method (CFM) to Determine the Null Parameters

When using the CFM one does not attempt to rotate the probe to the exact position where the yaw pressure is zero; instead, one measures each of the five physical variables shown in Table 3 over a narrow range of yaw pressures surrounding $\Delta P_{23} = 0$. Then, the measured values are fitted to a polynomial function of the yaw pressure, which is evaluated at $\Delta P_{23} = 0$ to give the respective null parameters (*e.g.*, θ_{NULL}, $\Delta P_{45,\text{NULL}}$, and $\Delta P_{12,\text{NULL}}$). We use the notation Γ_{q} to represent any of the five quantities listed in Table 3 where the subscript "q" identifies each quantity. The measured values are compactly expressed as ordered pairs ($\Delta P_{23,\text{n}}, \Gamma_{\text{q,n}}$), where the subscript *n* denotes the n$^{\text{th}}$ measurement of the *N* total measurements made at the points listed in Table 1 for the spherical probe and in Table 2 for the prism probe.

Table 3. *Definition of the Generic Probe Parameters* ($\Gamma_{\text{q,n}}$)

θ	Physical Variable	Generic Probe Parameters
1	yaw angle, (θ)	$\Gamma_1 = \theta$
2	pitch pressure, (ΔP_{45})	$\Gamma_2 = \Delta P_{45}$
3	velocity pressure, (ΔP_{12})	$\Gamma_3 = \Delta P_{12}$
4	pitch pressure ratio, (PR_1) [a]	$\Gamma_4 = PR_1 = \Delta P_{45}/\Delta P_{12}$
5	velocity pressure ratio, (PR_2) [b]	$\Gamma_5 = PR_2 = \sqrt{\Delta P_{\text{LDA}}/\Delta P_{12}}$

[a] By definition, the pitch pressure ratio equals the pitch calibration factor, $PR_1(0) = F_1$ at zero yaw pressure.
[b] By definition, the velocity pressure ratio equals the velocity calibration factor $PR_2(0) = F_2$ at zero yaw pressure.

The *N* data points ($\Delta P_{23,\text{n}}$, $\Gamma_{\text{q, n}}$) are fit to a 4$^{\text{th}}$ degree polynomial given by[8]

$$\Gamma_{\text{q,FIT}} = a_0 + a_1 \Delta P_{23} + a_2 \Delta P_{23}^2 + a_3 \Delta P_{23}^3 + a_4 \Delta P_{23}^4 \tag{7}$$

where a_0, a_1, a_2, a_3, and a_4, are the fit coefficients, and $\Gamma_{\text{q,FIT}}$ is the fitted function. The fit is used as a continuous extension of the discrete data set; therefore, the null parameters are determined by evaluating the curve fit at zero-yaw pressure,

8 At some large (positive and negative) pitch angles we used a 3$^{\text{rd}}$ degree polynomial.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$\Gamma_{q,FIT}(0) = \Gamma_{q,NULL} = a_{q,0} \tag{8}$$

where $a_{q,0}$ is the zeroth-degree fit coefficient (*i.e.*, *y*-intercept of the polynomial fit) the and the subscript "q" specifies the quantity that is being evaluated, as shown in Table 3. For example, for q = 1, 2, and 3 the corresponding null parameters are: $a_{1,0} = \theta_{NULL}$; $a_{2,0} = \Delta P_{45,NULL}$; and $a_{3,0} = \Delta P_{12,NULL}$, while for q = 4 and 5 the calibration coefficients are: $a_{4,0} = F_1$ and $a_{5,0} = F_2$.

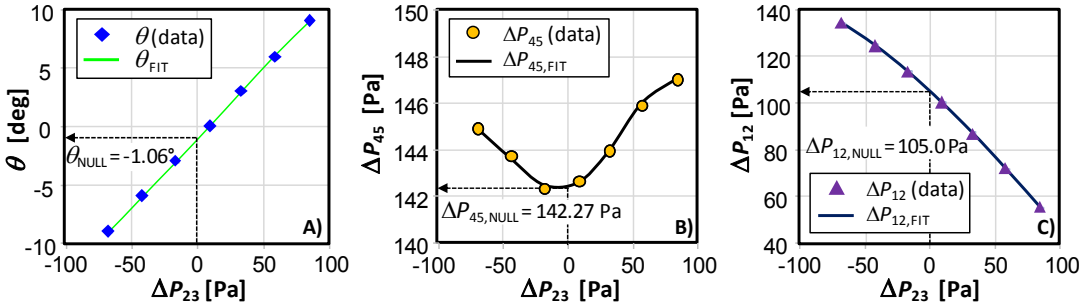


Figure 5. Plots of the A) the yaw angle (◆), B) the pitch pressure (●), and C) the velocity pressure (▲) versus the yaw angle for the spherical probe at an air speed of 15 m/s and a pitch angle of 21°. A 4th degree curve fit evaluated at $\Delta P_{23} = 0$ gives a yaw-null angle of $\theta_{NULL} = -1.06°$, a pitch pressure at the yaw-null angle of $\Delta P_{45,NULL} = 142.27$ Pa, and a velocity pressure at the yaw-null angle of $\Delta P_{12,NULL} = 105.0$ Pa.

Figure (5) shows an example of how the null parameters are computed for the spherical probe at $V_{LDA} = 15$ m/s and $\alpha = 21°$. The measured yaw angles (◆), pitch pressures (●), and velocity pressures (▲) are plotted along with their corresponding polynomial curve fits indicated by the solid lines (–). The dashed line (- - -) in each plot originates vertically from x-axis, intercepts the polynomial fit at $\Delta P_{23} = 0$, and extends horizontally to the y-axis where it indicates the value of respective null parameter. For the air speed and pitch angle in this example, Fig. (5A) shows the yaw-null angle equals $\theta_{NULL} = -1.06°$. Fig. (5B) shows the pitch pressure at the null angle is $\Delta P_{45,NULL} = 142.27$ Pa, and Fig. (5C) depicts that the velocity pressure at the null angle is $\Delta P_{12,NULL} = 105.0$ Pa. The measured null parameters along with the air speed and air density are used to calculate F_1 and F_2 and *via* Eqs. (6A) and (6B). For this example, these values are $F_1 = 1.358$ and $F_2 = 1.121$, respectively.

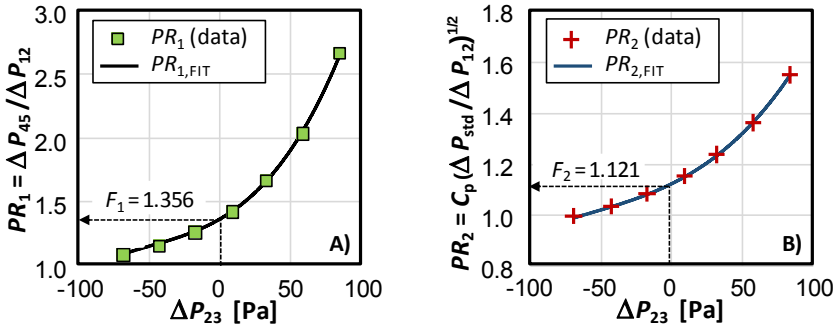


Figure 6. Plots of the A) the pitch pressure ratio (■) and B) the velocity pressure ratio (+) versus the yaw pressure for the spherical probe at an air speed of 15 m/s and a pitch angle of 21°. A 4th degree curve fit evaluated at $\Delta P_{23} = 0$ gives the pitch calibration factor of $F_1 = 1.356$ and a velocity calibration factor of $F_2 = 1.121$.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

As indicated in the last two rows of Table 3, the calibration factors F_1 and F_2 can be directly calculated using the CFM to evaluate pitch pressure ratio (PR_1) and the velocity pressure ratio (PR_2) at $\Delta P_{23} = 0$. Figure (6) illustrates how this is done. The pitch calibration factor (F_1) is determined by curve fitting PR_1 (■) versus the yaw pressure in Fig. (6A) and evaluating the fit at $\Delta P_{23} = 0$. Similarly, the velocity calibration factor (F_2) is the *y*-intercept of the curve fit of PR_2 (+) versus the yaw pressure shown in Fig. (6B). The difference in the values of F_1 and F_2 calculated by evaluating PR_1 and PR_2 at $\Delta P_{23} = 0$ shown in Fig. (6) are metrologically equivalent (*i.e.*, within the uncertainty) with F_1 and F_2 computed using the values of $\Delta P_{45,NULL}$ in Fig. (5B) and $P_{12,NULL}$ in Fig. (5C) along with Eqs. (6A) and (6B). Although both methods give equivalent results, in this work we opt to calculate the null parameters using Eqs. (6A) and (6B) as this choice facilitates a slightly easier uncertainty analysis.

4.3 Calibration Results for a Spherical Probe

Figure 7 shows the calibration data of the spherical probe performed on 7/12/2016. Figures 7A and 7B show F_1 and F_2, respectively, plotted as a function of the pitch angle α. Each symbol in the figure corresponds to one of the 6 air speed set points (V_n's) shown in the legend. Each of these velocities is measured by the LDA system. In Fig. 7A, the pitch angle calibration factor F_1 is approximately a linear function of α at pitch angles above –30°. In Fig. 7B, the values of F_2 are approximately a parabolic function of α. Although Figs. 7A and 7B display the trends of F_1 and F_2, they do not reveal the details of the velocity-dependences of F_1 and F_2.

The velocity-dependences of F_1 and F_2 are shown more clearly in Figs. 7C and 7D, respectively. Figure 7C plots the difference $F_1(V_n) - F_1(20\text{ m/s})$, where $F_1(V_n)$ is evaluated at air speeds V_n indicated by the legend above Fig. 7 and $F_1(20\text{ m/s})$ is the reference value at 20 m/s. Similarly, Fig. 7D is a plot of the percent difference $100\times[F_2(V_n)/F_2(20\text{ m/s}) - 1]$ between $F_2(V_n)$ and reference value $F_2(20\text{ m/s})$. Plotting the calibration factors relative to the reference values clearly shows that F_1 and F_2 are velocity-dependent. The velocity-dependence is largest at pitch angles $\alpha < -30°$ and $\alpha > 40°$. For large negative pitch angles, the values of $F_2(V_n)$ deviate by as much as 10 % from the reference value. However, for pitch angles $\alpha > -25°$ the velocity-dependence of $F_2(V_n)$ is less than 1 % of F_2, except for a fraction of the data at the lowest air speed $V_1 = 5$ m/s. These results show that one could calibrate a spherical probe at a single reference velocity of 20 m/s and at a single pitch angle $\alpha = 0°$, and then apply the calibration to a wider range of

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

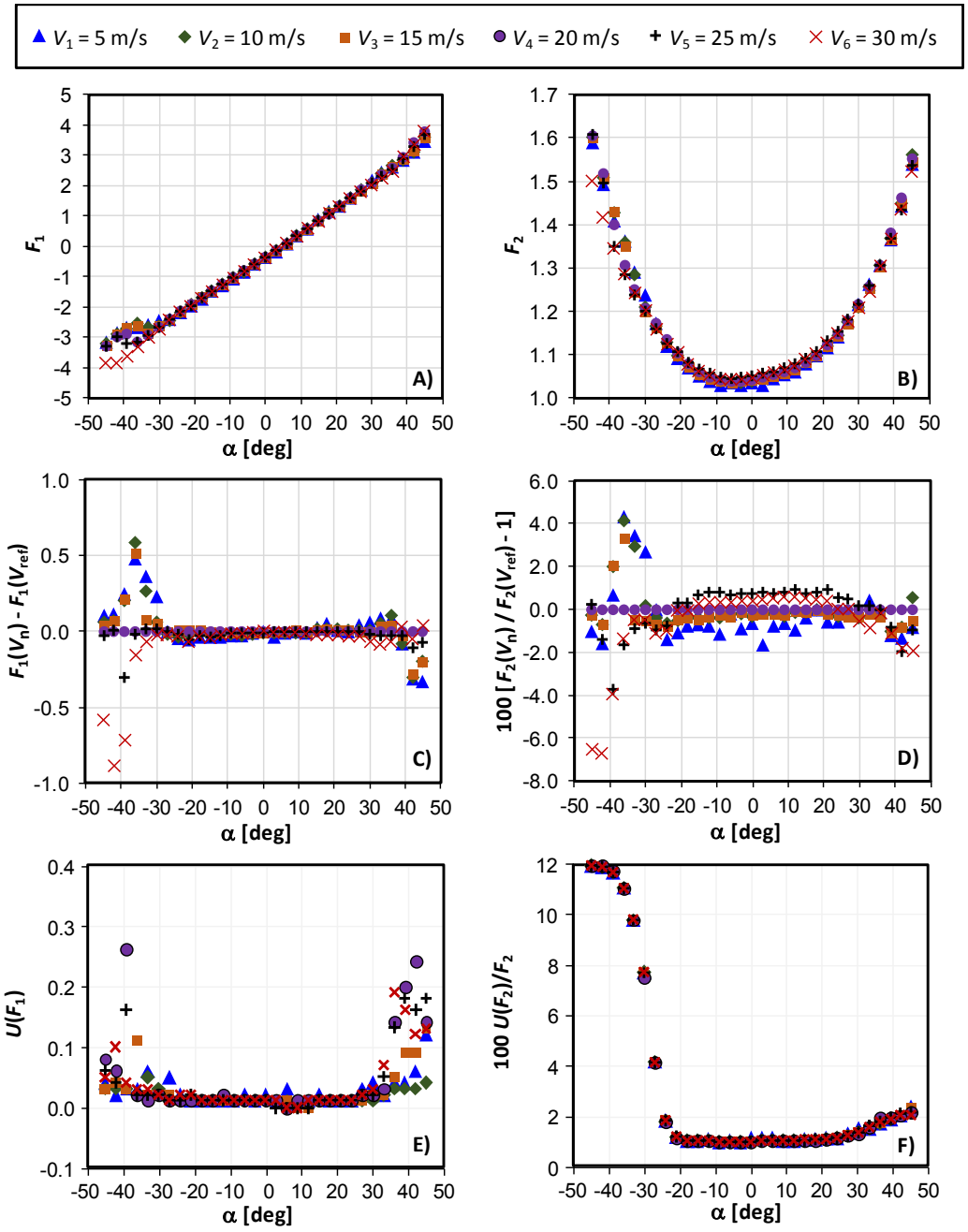


Figure 7. Calibration results of Spherical Probe for 6 air speeds ranging from 5 m/s to 30 m/s as functions of pitch angle α. A) pitch calibration factor, F_1; B) velocity calibration factor, F_2; C) difference between F_1 at specified air speeds and a reference value at an air speed of 20 m/s; D) percentage difference between F_2 at specified air speeds and a reference value at an air speed at 20 m/s; E) expanded uncertainty of F_1 at 95 % confidence level; F) expanded percentage uncertainty of F_2 at 95 % confidence level.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

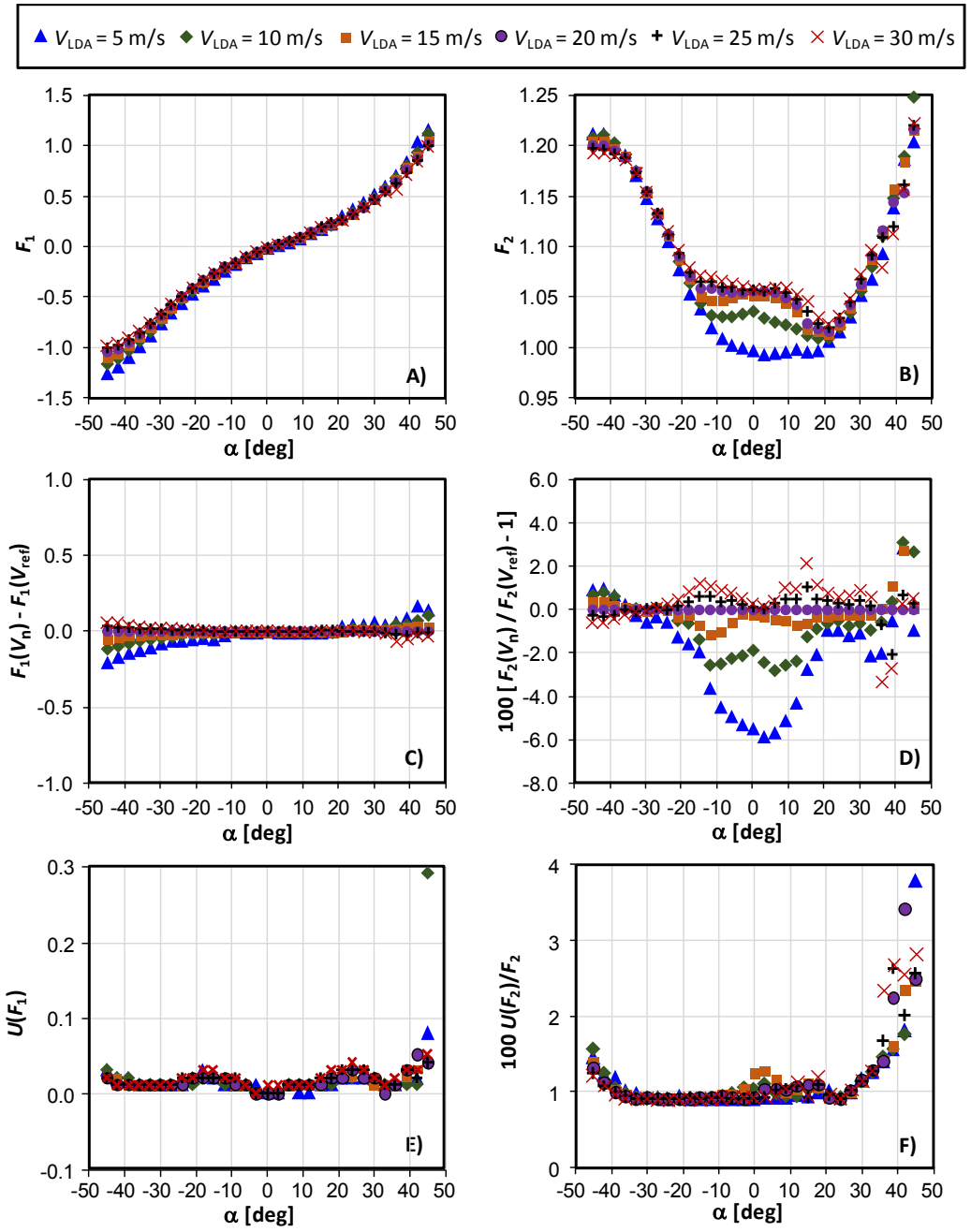


Figure 8. Calibration results of Prism Probe for 6 air speeds ranging from 5 m/s to 30 m/s as functions of pitch angle α. A) pitch calibration factor, F_1; B) velocity calibration factor, F_2; C) difference between F_1 at specified air speeds and F_1 at the reference air speed of 20 m/s; D) percentage difference between F_2 at specified air speeds and F_2 at the reference air speed at 20 m/s; E) expanded uncertainty of F_1 at 95 % confidence level; F) expanded percentage uncertainty of F_2 at 95 % confidence level.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

airspeeds (10 m/s $\le V_{LDA} \le$ 30 m/s) and pitch angles ($-25° \le \alpha \le 35°$). The values of F_2 determined from such a single-point calibration would be within ± 1 % of the values of F_2, as determined via the complete calibration.[9]

Figure 7E shows the expanded uncertainty $U(F_1)$ of the calibration factor F_1 and Fig 7F shows the percent expanded uncertainty $100 \times U(F_2)/F_2$. The expanded uncertainty is defined as the confidence interval that includes 95 % of the F_1 and F_2 data. The uncertainties $U(F_1)$ and $U(F_2)$ have two sources: 1) the uncertainty attributed to the reproducibility of repeated calibrations, and 2) the *data reduction uncertainty*, which combines uncertainty contributions from all the auxiliary variables used to calculate F_1 and F_2. Section 4.5 documents the reproducibility uncertainties, and Section 5 documents the data reduction uncertainties.

The uncertainties of both calibration factors increased at the extreme positive and negative values of pitch. The sharp increase of $U(F_2)$ at pitch angles below $\alpha < -30°$ is primarily attributed to the random uncertainty, which is calculated by the standard deviation of repeated measurements. We speculate that this large random uncertainty results from complex flow patterns over the temperature probe and its protective cylindrical tip (See Fig. 1B.). At large negative pitch angles, the vortices shed from these obstacles are likely to interfere with the differential pressure measurements between the ports, thereby resulting in a large random component of uncertainty.

4.4 Calibration Results for a Prism Probe

Figure 8 shows the calibration data of the prism probe performed on 7/18/2016. The calibration factors F_1 and F_2 are plotted as a function of the pitch angle α in Figs. 8A and 8B, respectively. Each plotted symbol corresponds to one of the 6 velocity set points V_n shown in the legend. In contrast to the spherical probe, the prism probe's F_2 calibration factor has a strong *Re*-dependence at pitch angles in the range $-20° \le \alpha \le 20°$. As shown in Fig. 8D, the values of $F_2(V_n)$ span a range of 6 %. [Fig. 8D plots the percentage difference between the velocity calibration factor at different air speeds V_n and a reference value at V_{ref} = 20 m/s, $100 \times [F_2(V_n)/F_2(20 \text{ m/s}) - 1]$. Analogously, Fig. 8C is a plot of the differences $F_1(V_n) - F_1(20 \text{ m/s})$. As shown in Fig. 8C, F_1 is independent of velocity for pitch angles in the range $-10° \le \alpha \le 20°$; however, outside this range F_1 becomes increasingly velocity dependent. Because of the prism probe's velocity dependence, it should be calibrated over the range of air speeds it will be used in stack measurement applications to achieve better accuracy.

The expanded uncertainties of the calibration coefficients $U(F_1)$ and $100 \times U(F_2)/F_2$ and are plotted in Figs. 8E and 8F. The uncertainty of F_1 is nearly independent of pitch angles in the range $-40° \le \alpha \le 40°$; it increases when $\alpha > 40°$. As shown in Fig. 8F, the expanded uncertainty $100 \times U(F_2)/F_2$ increases sharply at large positive ($\alpha > 40°$) and negative ($\alpha < -40°$) pitch angles. For pitch angles $-30° \le \alpha \le 0°$ the expanded uncertainty is approximately a constant value of 0.9 %. In the range $0° \le \alpha \le 20°$, the expanded uncertainty increases only slightly, reaching a maximum value of 1.25 %.

4.5 Probe Reproducibility and Rational Polynomial Curve Fits of the Calibration Data

During a EPA stack emissions testing commonly called a relative accuracy test audit (RATA), the axial velocity along the stack axis ($V_{z'}$) is determined by the equation:

9 While this observation applies for the laminar flow conditions used in this calibration, more research is needed to confirm it applies in turbulent flow.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$V_x = F_2\sqrt{\frac{2\Delta P_{12,\mathrm{NULL}}}{\rho}}\cos(\theta_{\mathrm{NULL}})\cos(\alpha) \tag{9}$$

where the velocity pressure $\Delta P_{12,\mathrm{NULL}}$ and the yaw-null angle θ_{NULL} are measured using the Method 2F nulling procedure. The pitch pressure is also measured at the null condition $\Delta P_{45,\mathrm{NULL}}$ and used along with $\Delta P_{12,\mathrm{NULL}}$ to calculate the pitch calibration factor F_1 *via* Eq. (6A). In contrast, the pitch angle α and velocity calibration factor F_2 are not measured during a RATA test, but instead are based on curve fits of the probe calibration data. For a specified probe shape, the calibration factors F_1 and F_2 are characterized by the pitch angle, Reynolds number (or airspeed), and turbulence intensity.[10] For some probe shapes the *Re*-dependence is significant (*e.g.*, see prism probe results for F_2 in Fig. 8D) while for other probes the *Re*-dependence is small (*e.g.*, see spherical probe results for F_2 in Fig. 7D).

Here, we provide high-accuracy curve fits to the pitch angle and the velocity calibration factor. These fits account for Reynolds number effects and are valid for airspeeds spanning the range 5 m/s $\le V \le$ 30 m/s and for pitch angles $-45° \le \alpha \le 45°$. We use a rational polynomial fit expressed by

$$\Psi_{\mathrm{FIT,p}} = \sum_{m=0}^{N}\sum_{n=0}^{N-m} a_{\mathrm{mn,p}} x_{\mathrm{p}}^{\mathrm{m}} y_{\mathrm{p}}^{\mathrm{n}} \Bigg/ \sum_{m=0}^{N}\sum_{n=0}^{N-m} b_{\mathrm{mn,p}} x_{\mathrm{p}}^{\mathrm{m}} y_{\mathrm{p}}^{\mathrm{n}} \tag{10}$$

where the numerator and denominator are each N = 5 degree polynomials of the two variables x_{p} and y_{p}. The fit was developed using the commercial available software package Orgin.[6] The index "p" ranges from 1 to 3; it identifies the fitting functions $\Psi_{\mathrm{FIT,p}}$, and their dependent variables x_{p} and y_{p}, as shown in Table 4. Using this compact notation, the symbol $\Psi_{\mathrm{FIT,p}}$ represents three fitted functions; 1) $\Psi_{\mathrm{FIT,1}} = \alpha_{\mathrm{FIT}}$ is the pitch angle fitted to the dependent variables $\Delta P_{45,\mathrm{NULL}}$ and F_1; 2) $\Psi_{\mathrm{FIT,2}} = F_{2,\mathrm{FIT}}$ is the velocity calibration factor fitted to the dependent variables α_{FIT} and $(\Delta P_{12,\mathrm{NULL}})^{1/2}$, and 3) $\Psi_{\mathrm{FIT,3}} = \Delta P_{\mathrm{std,FIT}}$ is the standard pressure fitted to the dependent variables $(\Delta P_{12,\mathrm{NULL}})^{1/2}$ and F_1. The third function $\Psi_{\mathrm{FIT,3}}$ can be used instead of $\Psi_{\mathrm{FIT,2}}$ to determine the flow velocity. In this case the grouped-terms $F_2^2\Delta P_{12,\mathrm{NULL}}$ in Eq. (9) are replaced by ΔP_{std} so that the axial velocity is expressed in terms of the standard pressure, $V_x = \sqrt{2\Delta P_{\mathrm{std}}/\rho}\cos(\theta_{\mathrm{NULL}})\cos(\alpha)$.

Table 4. *Rational polynomial fits for α, F_2, and ΔP_{std}.*

p	x_{p}	y_{p}	$\Psi_{\mathrm{FIT,p}}$
1	$\Delta P_{45,\mathrm{NULL}}$, [Pa]	F_1, [ ]	α_{FIT}, [deg]
2	α_{FIT}, [deg]	$\sqrt{\Delta P_{12,\mathrm{NULL}}}$, [Pa$^{1/2}$]	$F_{2,\mathrm{FIT}}$, [ ]
3	$\sqrt{\Delta P_{12,\mathrm{NULL}}}$, [Pa$^{1/2}$]	F_1, [ ]	$\Delta P_{\mathrm{std,FIT}}$, [Pa]

10 The effect of turbulence intensity on the calibration factors is a topic of current research and is beyond the scope of this document. Preliminary studies show that a 10 % turbulence intensity has a 2 % - 3 % effect on the velocity calibration factor, F_2.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

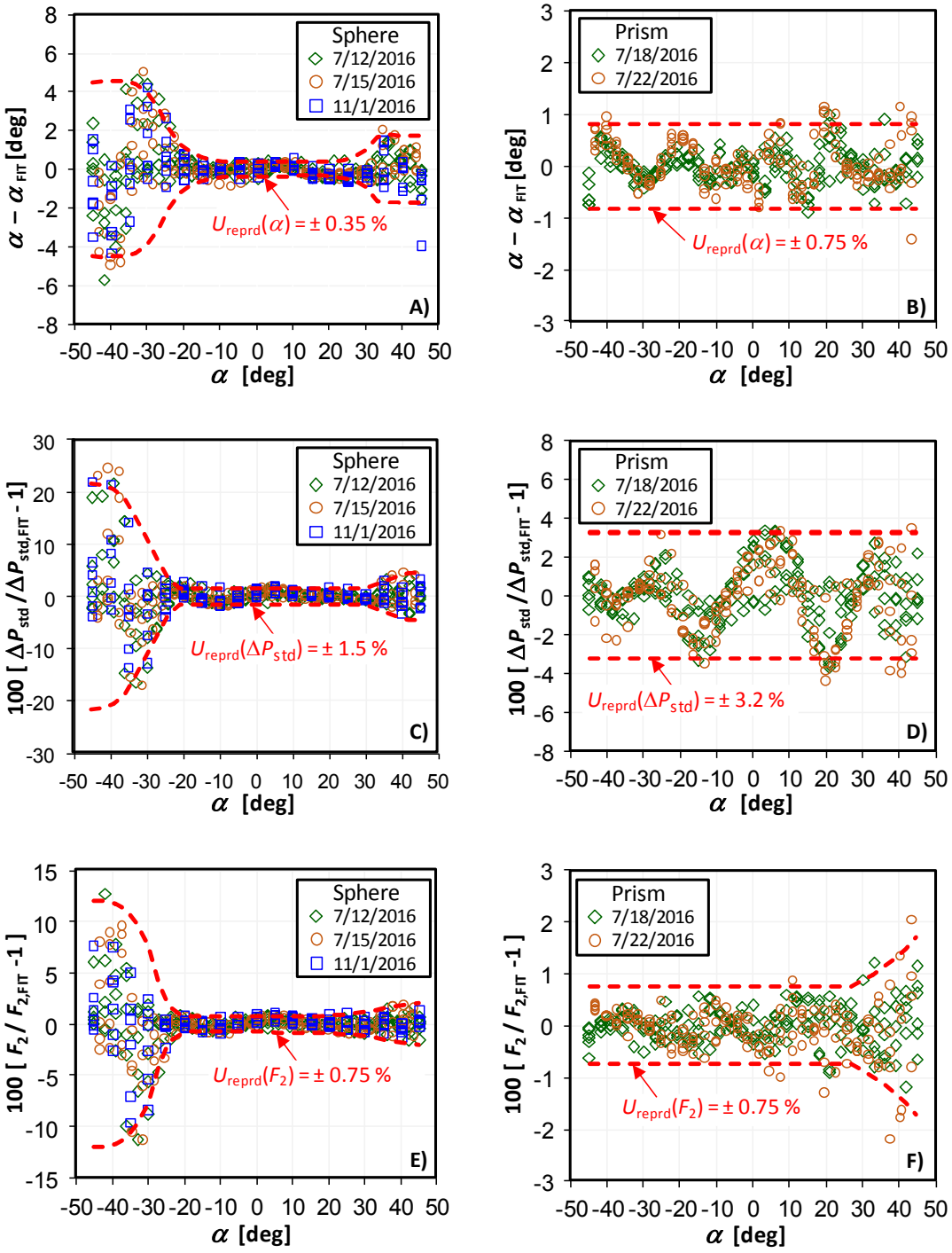


Figure 9. Residuals from Eqs. (10), rational polynomials fitted to 3 repeated calibrations of the spherical probe (left) and the 2 repeated calibrations of the prism probe (right). The residuals are plotted versus the pitch angle α for 6 air speeds ranging from 5 m/s to 30 m/s. The dashed lines (- - -) enclose 95 % of the data.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

The residuals of the three fitted functions $\Psi_{FIT,p}$ from the corresponding measured values are plotted as a function of pitch angle in Fig. 9. Figures. 9A, 9C, and 9E are plots of the residuals from three repeated calibrations of the spherical probe; Figs. 9B, 9D, and 9F are plots of the residuals from two calibrations of the prism probe. For comparisons, the residuals from the two probes are plotted next to each other. (Note: the vertical scales for the spherical probe span wider ranges than the corresponding scales for the prism probe.) Figures 9A and 9B plot the pitch angle residual $\alpha - \alpha_{FIT}$. Figures 9C and 9D plot as a percent, the residual of the fitted standard pressure from the measured pressure $100\times[\Delta P_{std}/\Delta P_{std,FIT} - 1]$. Figures 9E and 9F plot as a percent, the residual of the fitted velocity factor from the measured velocity factor $100\times[\Delta F_2/\Delta F_{2,FIT} - 1]$.

The dashed lines (- - -) are confidence intervals containing 95 % of the data. If data in the confidence intervals were randomly distributed, these confidence intervals would be a measure of the reproducibility U_{reprd} of repeated calibrations. In Fig. 9D, the averaged data at some of the pitch angles is not zero, indicating that the deviations are not completely random. A better curve fit could reduce these deviations; thereby, reducing U_{reprd}. Thus, the structure in the residual plots cause the values of U_{reprd} to be biased high. Nevertheless, we use this measure to estimate the reproducibility of the calibration factors, and the uncertainties attributed to reproducibility of F_1 and F_2 are included in the respective uncertainty budgets for the spherical probe in Section 4.3 and the prism probe in Section 4.4.

Care must be taken when comparing residual plots of the spherical probe to those of the prism probe because the scales on the y-axes are not the same. However, it is clear all the residual plots of the spherical probe sharply increase at large negative pitch values. In Section 4.3 we hypothesized that at large negative pitch angles the pressure ports are engulfed by a turbulent wake. The wake is caused by vortices shed from the temperature sensor and the cylindrical tip located just upstream of the probe ports (see probe geometry in Fig. 1B) at large negative pitch angles. The inherent fluctuations in the wake result in the larger reproducibility for $\alpha < -20°$. In field RATA tests, pitch angles are usually greater than $-20°$; therefore, this reproducibility will not limit the accuracy of RATA tests using spherical probes.

For pitch angles $-20° \le \alpha \le 30°$ the values of U_{reprd} for the spherical probe (See Figs. 9A, 9C, and 9E.) are uniform and are less than or equal to U_{reprd} of the prism probe. This range of pitch angles includes typical α values encountered in typical RATA tests; therefore, we expect slightly better reproducibility from a spherical probe than from a prism probe.

At large positive pitch angles ($\alpha > 30$), the reproducibility of the spherical probe is not quite as good, as evidenced by the slight increase in U_{reprd} in probe in Figs. 9A, 9C, and 9E. Figure 9F shows that the reproducibility of the velocity calibration factor for the prism probe also increases at large positive pitch angles. The reproducibility at large positive pitch angles, like that of large negative pitch angles, does not play a role in field RATA tests where pitch angles are typically $-20° \le \alpha \le 20°$.

5 UNCERTAINTY ANALYSIS OF CALIBRATION RESULTS

In this section we calculate the expanded uncertainty of the yaw-null angle θ_{NULL}, the pitch pressure at the null condition $\Delta P_{45,NULL}$, the velocity pressure at the null condition $\Delta P_{12,NULL}$, and the calibration factors F_1 and F_2. These uncertainties are calculated for airspeeds $5 \text{ m/s} \le V_{LDA} \le 30 \text{ m/s}$ and pitch angles from $-45 \le \alpha \le 45$. The uncertainties of the pitch (F_1) and velocity (F_2) calibration factors are plotted in Figs. 7E and 7F for the spherical probe, and in Figs. 8E and 8F for the prism probe. In most stack applications, the maximum range of the pitch

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

angle is $-20 \le \alpha \le 20$. For this narrower range of pitch angles the expanded uncertainty of F_2 is nearly constant for both probes; for the spherical probe, $U_r(F_2)$ ranges from 0.9 % to 1.1 % while for the prism probe, $U_r(F_2)$ ranges from 0.9 % to 1.25 %. For both probes the largest uncertainty sources for F_2 are the reproducibility of repeated calibrations and the LDA standard, which together, contributed 60 % to 80 %.

The methodical uncertainty calculations provided herein are intended both to document these results and to provide guidance to those in stack testing community or other metrologist not as familiar with uncertainty analysis. Those not interested in these details can skip this section without loss of continuity.

5.1 Data Reduction Equation and Method of Propagation of Uncertainty

The uncertainty analysis is based on the method of propagation of uncertainty [13,14]. In this method, a mathematical expression called the *data reduction equation* relates the functional dependence of one or more measured input quantities to an output quantity. Equation (11) is a generic data reduction equation

$$y = y(x_1, x_2, \ldots, x_M), \tag{11}$$

that relates M input quantities given by the x_m's to a single output y. Several data reduction equations have been used throughout this work; for example, the air density ρ in Eq. (2) is not measured directly, but is calculated from its data reduction equation, which has as its inputs the measured pressure P, temperature T, and relative humidity RH. Likewise, the pitch calibration factor F_1 is calculated from its data reduction equation given in Eq. (6A), which has as its inputs $\Delta P_{45,NULL}$ and $\Delta P_{12,NULL}$.

Propagated Uncertainties

The same data reduction equation that computes the value of an output quantity as a function of known input quantities is also used to relate the uncertainty of the M input quantities, $u(x_m)$'s, to the uncertainty of the output quantity, $u_{c,p}(y)$, called the *standard combined, propagated uncertainty*. That is, the uncertainty of an output quantity is determined by propagating the known uncertainties of the input quantities through the corresponding data reduction equation. For the generic data reduction equation given by Eq. (11) the standard combined, propagated uncertainty is

$$u_{c,p}(y) = \sqrt{\sum_{m=1}^{M} \left(\frac{\partial y}{\partial x_m} \right)^2 u^2(x_m) + \sum_{m=1}^{M} \sum_{n=1}^{M} (1 - \delta_{mn}) \left(\frac{\partial y}{\partial x_m} \right) \left(\frac{\partial y}{\partial x_n} \right) r(x_m, x_n) u(x_m) u(x_n)} \tag{12A}$$

where $u_{c,p}(y)$ is the uncertainty at the 68 % confidence level, the $u(x_m)$'s are the standard uncertainties of the input quantities at the 68 % confidence level; the $\partial y/\partial x_m$'s are the dimensional sensitivity coefficients equal to the partial derivatives of the output variable (y) with respect to each input variable (x_m); δ_{mn} is the *Kronecker delta function*, which equals 1 if m = n and 0 if m ≠ n; and $r(x_m, x_n)$ is the normalized correlation function, which indicates the degree of correlation between x_n and x_m.

The value of the normalized correlation coefficient in Eq. (12A) has the range $-1 \le r(x_m, x_n) \le 1$. In many cases, its value is difficult to determine, and approximations are necessary. In the special case where all the uncertainty sources are uncorrelated $r(x_m, x_n) = 0$, and the standard combined, propagated uncertainty simplifies to

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$u_{c,p}(y) = \sqrt{\sum_{m=1}^{M} \left(\frac{\partial y}{\partial x_m} \right)^2 u^2(x_m)} \,, \tag{12B}$$

which is the root-sum-square (RSS) of the standard uncertainty of each input amplified by its sensitivity coefficient. Alternatively, if the uncertainty sources are perfectly correlated then $r(x_m, x_n) = 1$, and Eq. (12A) simplifies to

$$u_p(y) = \left| \sum_{n=1}^{N} \left(\frac{\partial y}{\partial x_n} \right) u(x_n) \right| \tag{12C}$$

where the propagated uncertainty equals the absolute value of the summed products of $u(x_n)$ and $\partial y / \partial x_n$.

Throughout this document most of the propagated uncertainty sources are uncorrelated, and the RSS formulation in Eq.(12B) is used. However, in Section 5.2 the uncertainty of the null parameters has partially correlated uncertainty sources and $r(x_m, x_n)$ cannot be taken to be 0 or 1, but has an intermediate value between these two limits. As such, in this section we use an approximate method similar to Coleman [14] to estimate the degree of correlation. The details are explained in Section 5.2.

Non-Propagated Uncertainties

The uncertainty in the measuraund y does not derive exclusively from propagation of measurement errors in the input quantities. Herein, we use the notation $u_{c,np}(y)$ to denote the *standard combined, non-propagated uncertainties*. In this document, non-propagated uncertainties include 1) the measurement reproducibility of F_1 and F_2, and 2) the null-parameter curve fit uncertainty.

Combined and Expanded Uncertainties

The *standard combined uncertainty* is the RSS of the standard combined, propagated uncertainty, $u_{c,p}(y)$, and the standard combined, non-propagated uncertainty, $u_{c,np}(y)$, as given by

$$u_c(y) = \sqrt{u_{c,p}^2(y) + u_{c,np}^2(y)} \,, \tag{13A}$$

and the expanded uncertainty is

$$U(y) = k u_c(y) = k \sqrt{u_{c,p}^2(y) + u_{c,n}^2(y)} \tag{13B}$$

where $u_c(y)$ is the standard combined uncertainty, and k is the coverage factor, which when taken equal to $k = 2$, converts the combined standard uncertainty at the 68 % confidence level to the expanded uncertainty $U(y)$ at the 95 % confidence level, ($U = k\, u_c$).

5.2 Uncertainty of the Null Parameters $u(\Delta P_{23,\,NULL})$, $u(\Delta P_{45,NULL})$, and $u(\Delta P_{12,NULL})$

Using the CFM, the null parameters are calculated numerically via a linear regression algorithm. As shown in Eq. (8) these parameters are equal to the regression coefficient $a_{q,0}$. As such, the functional dependence of $a_{q,0}$ on its input parameters is not explicitly known. In contrast, we point out that the functional dependence of the F_1 and F_2 calibration coefficients are explicitly defined

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

by Eqs. (6A) and (6B), respectively. Since $a_{q,0}$ cannot be characterized by such simple algebraic expressions, we represent the output of the linear regression algorithm by the following generic expression

$$a_{q,0} = a_{q,0}(\Delta P_{23_1},\ldots,\Delta P_{23_n},\ldots,\Delta P_{23_N},\Gamma_{q,1},\ldots,\Gamma_{q,n},\ldots,\Gamma_{q,N}), \quad (14)$$

where N is the number of data points used in the curve fit as specified in Tables 1 and 2; $\Delta P_{23,n}$ is the nth measurement of the yaw pressure; and $\Gamma_{q,n}$ is the nth measurement of the generic probe variable defined in Table 3 of Section 4.2. The dependent variables on the right-hand side of Eq. (14) list the N data points, ($\Delta P_{23,n}$, $\Gamma_{q,n}$), used by the regression model to calculate $a_{q,0}$. Because $a_{q,0}$ is calculated numerically, the sensitivity coefficients must also be calculated numerically. This is accomplished using a finite difference technique whereby we perturb a single dependent variable in Eq. (14) by a small amount ε, re-compute the new perturbed regression coefficients via the least squares method, subtract the new coefficient $\hat{a}_{q,0}$ from the unperturbed value $a_{q,0}$, and divide the quantity by ε [15]. The complete set of sensitivity coefficients are determined by repeating this procedure for each one of the dependent variables.

Non-Propagated Uncertainty

The non-propagated uncertainty is given by

$$u_{np}(a_{q,0}) = \sqrt{\frac{\sigma_{FIT}^2}{3} + \sigma_{RES}^2 \sum_{n=1}^{N}\left(\frac{\partial a_{q,0}}{\partial \Gamma_{q,n}}\right)^2} \quad (15A)$$

where $\sigma_{FIT}/\sqrt{3}$ is the standard deviation of the mean of 3 values of $a_{q,0}$ calculated using 2nd, 3rd, and 4th degree polynomial fits. This uncertainty accounts for the fact that there is no physical basis to select the degree of the polynomial to use. The second term is the "so called" classical solution for the uncertainty of the y-intercept [14]. It is calculated by multiplying the sensitivity coefficient by the standard deviation of the curve fit residuals or also known as the standard error of regression (σ_{RES}), which is calculated by [16]

$$\sigma_{RES} = \sqrt{\frac{1}{N-2}\sum_{n=1}^{N}\left(\Gamma_{q,FIT}(\Delta P_{23_n}) - \Gamma_{q,n}\right)^2}. \quad (15B)$$

The second term accounts for the uncertainty of the curve fit, but does not include systematic uncertainties attributed to the zero-pressure offset, calibration of the pressure sensors, etc. These uncertainties are accounted for in the propagated uncertainty.

Propagated Uncertainty

Measurement errors in $\Delta P_{23,n}$ and in $\Gamma_{q,n}$ propagate through Eq. (14) and lead to the following expression for propagated uncertainty,

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$
\begin{aligned}
u_{\mathrm{p}}^2(a_{\mathrm{q},0}) = & \sum_{n=1}^{N}\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Delta P_{23,\mathrm{n}}}\right)^2 u^2(\Delta P_{23_{\mathrm{n}}}) \\
& + \sum_{m=1}^{N}\sum_{n=1}^{N}(1-\delta_{\mathrm{mn}})\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Delta P_{23,\mathrm{m}}}\right)\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Delta P_{23,\mathrm{n}}}\right) r(\Delta P_{23_{\mathrm{m}}}, \Delta P_{23,\mathrm{n}}) u(\Delta P_{23,\mathrm{m}}) u(\Delta P_{23,\mathrm{n}}) \\
& + \sum_{n=1}^{N}\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Gamma_{\mathrm{q,n}}}\right)^2 u^2(\Gamma_{\mathrm{q,n}}) + \sum_{m=1}^{N}\sum_{n=1}^{N}(1-\delta_{\mathrm{mn}})\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Gamma_{\mathrm{q,m}}}\right)\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Gamma_{\mathrm{q,n}}}\right) r(\Gamma_{\mathrm{q,m}}, \Gamma_{\mathrm{q,n}}) u(\Gamma_{\mathrm{q,m}}) u(\Gamma_{\mathrm{q,n}}) \\
& + \sum_{m=1}^{N}\sum_{n=1}^{N}(1-\delta_{\mathrm{mn}})\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Delta P_{23,\mathrm{m}}}\right)\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Gamma_{\mathrm{q,n}}}\right) r(\Delta P_{23,\mathrm{m}}, \Gamma_{\mathrm{q,n}}) u(\Delta P_{23,\mathrm{m}}) u(\Gamma_{\mathrm{q,n}})
\end{aligned}
\quad . \qquad (16A)
$$

Generally speaking, the input variables are correlated so that the correlation coefficients $r(\Delta P_{23,\mathrm{m}}, \Delta P_{23,\mathrm{n}})$, $r(\Gamma_{\mathrm{q,m}}, \Gamma_{\mathrm{q,n}})$, and $r(\Delta P_{23,\mathrm{m}}, \Gamma_{\mathrm{q,n}})$ are not zero, but have some intermediate value between 0 and 1. We estimate the values of the correlation coefficient using an approach similar to Coleman [14]. Each correlation coefficient and its corresponding uncertainties are decomposed into their elementary or base uncertainty sources. Elemental uncertainty sources include: 1) the calibration uncertainty of the pressure transducers, $u(\Delta P_{\mathrm{cal}})$; 2) the zero-drift of pressure transducers, $u(\Delta P_{\mathrm{zero}})$; and 3) the uncertainty of yaw angle alignment, $u(\theta)$. The elemental uncertainty sources have cross-correlation coefficients that are equal to zero; however, the elemental uncertainty sources may or may not be correlated with themselves. For example, since all the transducers are calibrated by the same pressure standard the calibration uncertainty is taken to be perfectly correlated when the sensors are used to measure the same nominal pressure. Thus, at the yaw-null pressure condition of $\Delta P_{23} = 0$, the calibration uncertainties of the two pressure transducers that measure ΔP_{12} and ΔP_{13} are perfectly correlated, and the correlation coefficient is unity, $r(\Delta P_{13,\mathrm{cal}}, \Delta P_{12,\mathrm{cal},}) = 1$. On the other hand, the calibration uncertainties are taken to be uncorrelated and have correlation coefficient equal to zero for the same transducer (or different transducers) used to measure different pressures. In cases where it is unclear whether an elemental correlation coefficient should be 0 or 1, we conservatively use the value that results in the larger uncertainty. By associating the perfectly correlated elemental uncertainties with Eq. (12B) and uncorrelated uncertainties with Eq. (12C) we simplify the propagated uncertainty in Eq. (16A) to

$$
u_{\mathrm{p}}(a_{\mathrm{q},0}) = \sqrt{\left[\sum_{n=1}^{N}\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Delta P_{23_{\mathrm{n}}}}\right) u(\Delta P_{23_{\mathrm{n}}})\right]^2 + \left[\sum_{n=1}^{N}\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Gamma_{\mathrm{q,n}}}\right) u(\Gamma_{\mathrm{q,1,n}})\right]^2 + \left[\sum_{n=1}^{N}\left(\frac{\partial a_{\mathrm{q},0}}{\partial \Gamma_{\mathrm{q,n}}}\right) u(\Gamma_{\mathrm{q,2,n}})\right]^2} \; . \qquad (16B)
$$

where the term $u(\Delta P_{23,\mathrm{n}})$ is the uncertainty in the yaw pressure measurement given in Eq. (4B), and $u(\Gamma_{\mathrm{q,1,n}})$ and $u(\Gamma_{\mathrm{q,2,n}})$ are the uncertainties of the generic probe parameters given by

$$
u(\Gamma_{\mathrm{q,1,n}}) = \begin{cases} u(\theta) & q = 1 \\ \sqrt{2}u(\Delta P_{\mathrm{cal}}) & q = 2 \\ u(\Delta P_{\mathrm{cal}}) & q = 3 \end{cases}, \qquad (16C)
$$

and

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$$u(\Gamma_{q,2,n}) = \begin{cases} 0 & q = 1 \\ \sqrt{2}u(\Delta P_{zero}) & q = 2 \\ u(\Delta P_{zero}) & q = 3 \end{cases}, \tag{16D}$$

respectively. The use of this approximate method is justified since it yields a conservative estimate of the uncertainty, and $u_p(a_{q,0})$ is not the dominate source of uncertainty in the uncertainty budget.

Combined Uncertainty

The combined uncertainty, including both propagated and non-propagated components is

$$u_c(a_{q,0}) = \sqrt{\frac{\sigma_{FIT}^2}{3} + \sigma_{RES}^2 \sum_{n=1}^{N}\left(\frac{\partial a_{q,0}}{\partial \Gamma_{q,n}}\right)^2 + \left[\sum_{n=1}^{N}\left(\frac{\partial a_{q,0}}{\partial \Delta P_{23_n}}\right)u(\Delta P_{23_n})\right]^2 + \left[\sum_{n=1}^{N}\left(\frac{\partial a_{q,0}}{\partial \Gamma_{q,n}}\right)u(\Gamma_{q,1,n})\right]^2 + \left[\sum_{n=1}^{N}\left(\frac{\partial a_{q,0}}{\partial \Gamma_{q,n}}\right)u(\Gamma_{q,2,n})\right]^2}. \tag{17}$$

5.3 Uncertainty Analysis of F_1 and F_2.

The uncertainty of the calibration factors F_1 and F_2 in Eqs. (6A) and (6B) are determined using Eq. (12B). The expression for the expanded uncertainty expression of F_1 is

$$U(F_1) = 2\sqrt{\frac{u^2(\Delta P_{45,NULL})}{\Delta P_{12,NULL}^2} + \frac{u^2(\Delta P_{12,NULL})F_1^2}{\Delta P_{12,NULL}^2} + u_{reprd}^2(F_1)} \tag{18A}$$

where $u(\Delta P_{45,NULL})$ and $u(\Delta P_{12,NULL})$ are the uncertainties of the null parameters specified in Eq. (17) for q = 2 and 3, respectively. The values of the null parameters are determined using the CFM from Section 4.2. Also, $u_{reprd}(F_1)$ is the standard reproducibility uncertainty documented in Section 4.5.

The expression for the expanded uncertainty for F_2 is

$$U_r(F_2) = 2\sqrt{\frac{u_r^2(\rho)}{4} + u_r^2(V_{LDA}) + \frac{u_r^2(\Delta P_{12,NULL})}{4} + u_{reprd}^2(F_2)} \tag{18B}$$

where $u_r(V_{LDA})$ is the relative standard uncertainty of the LDA specified in Section 2, $u_r(\rho)$ is the relative standard uncertainty in density specified in Section 2, and $u_{repro}(F_2)$ is the relative standard uncertainty attributed to reproducibility specified in Section 4.5

The expanded uncertainties of F_1 and F_2 are plotted in Figs. 7E and 7F for the spherical probe and in Figs. 8E and 8F for the prism probe. The expanded uncertainty of F_2 was comparable for both probes for α = - 20 to 20 . The prism probe exhibited an uncertainty ranging from 0.9 % to 1.25 % depending on airspeed while the uncertainty of the spherical probe ranges from 0.9 % to 1.1 %. The largest uncertainty sources for both probes include the reproducibility of repeated calibrations and the LDA airspeed standard, which together, contributed between 60 % and 80 % in all cases.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

6 LIMITATIONS OF METHOD 2F CAN CAUSE LARGE UNCERTAINTIES IN FIELD MEASUREMENTS

We identified two limitations of Method 2F that can result in large biases when 3-D probes are used in field measurements. First, Method 2F does not adequately account for the *Re*-dependence of probe calibration factors; instead Method 2F approximates F_1 and F_2 as a function of α only.[11] This approach is reasonable for the spherical probe for pitch angles $-20 \leq \alpha \leq 20$ since $F_{2,\mathrm{SPHERICAL}}$ is nearly independent of Reynolds number (or air speed). In contrast, $F_{2,\mathrm{PRISM}}$ strongly depends on Reynolds number, especially at the lower air speeds < 10 m/s. If we had calibrated the prism probe using Method 2F and then used the probe in the field, we would over-report low velocity (5 m/s) measurements by nearly 6 %. Therefore, to obtain accurate low velocities measurements in field applications, a prism probe must be calibrated over the range of velocities it will be used.

Second, more stringent uncertainty requirements are needed for the yaw-null pressure measurement at low air speeds (< 10 m/s). Method 2F specifies the yaw pressure transducer should have full-scale range of ± 124.5 Pa and an expanded uncertainty of 2.5 Pa (1 % of full-scale). In addition, during stack application, the protocol allows the yaw pressure transducer to have a zero drift as large as 7.5 Pa. If these specifications are met, the uncertainty in yaw pressure $u(\Delta P_{\mathrm{yaw}})$ will significantly increase the measurement uncertainty at low flue gas velocities.

Even if $u(\Delta P_{\mathrm{yaw}})$ is negligible, a third uncertainty source, not considered in Method 2F, should be considered. This uncertainty $u(\Delta P_{23\neq 0})$ occurs when the null parameters (θ_{NULL}, $\Delta P_{45,\mathrm{NULL}}$, and $\Delta P_{12,\mathrm{NULL}}$) are measured at a non-zero yaw-pressure $\Delta P_{23} \neq 0$. In practice, noisy pressure signals (especially during in-stack applications) make it difficult to find the exact yaw-null angle where $\Delta P_{23} = 0$. Consequently, the null parameters are generally measured at pressures slightly offset from zero, thereby increasing the uncertainty.

Accounting for all three uncertainty sources, the total yaw-pressure uncertainty is

$$u(P_{\mathrm{offset}}) = \sqrt{u^2(\Delta P_{\mathrm{cal}}) + u^2(\Delta P_{\mathrm{zero}}) + u^2(\Delta P_{23\neq 0})} \tag{19}$$

where $u(\Delta P_{\mathrm{cal}})$ is the uncertainty of the yaw pressure transducer, $u(\Delta P_{\mathrm{zero}})$ is the uncertainty of the zero-drift during a measurement, and $u(\Delta P_{23\neq 0}) = \Delta P_{23} \big/ \sqrt{3}$ is the uncertainty in yaw-pressure because it is slightly offset from zero, $\Delta P_{23} \neq 0$. We take the latter uncertainty to have rectangular distribution so that ΔP_{23} is multiplied by the factor $1/3^{1/2}$ [12].

We emphasize that the Method 2F allowable-yaw-pressure, zero-offset uncertainty will be a significant contributor to the overall uncertainty budget at low flows because $u(P_{\mathrm{offset}})$ will be a significant fraction of the dynamic pressure.

In our current work, $u(P_{\mathrm{offset}})$ is negligible relative to other uncertainty sources. First, we used the CFM introduced in Section 4.2 to ensure that the null parameters are determined at exactly

11 Method 2F requires that the calibration factors are measured over a range of pitch angles, but at each pitch angle F_1 and F_2 are only measured at 2 air speeds consisting of 18.3 m/s and 27.4 m/s. The average F_1 and F_2 at these 2 air speeds are curve fit as a function of α, and the calibration curve fits are extrapolated to lower flue gas velocities in stack applications [5].

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

$\Delta P_{23} = 0$. Second, we used differential pressure transducers (described in Section 3.1) with a small zero-drift uncertainty relative to the dynamic pressure of the air stream specified in Eq. (1). At the lowest velocity setting of $V_{LDA} = 5$ m/s the dynamic pressure is $\Delta P_{dyn} = 15$ Pa so that $u(\Delta P_{offset}) / \Delta P_{LDA} < 0.001$. In contrast, Method 2F specifications for the uncertainty of yaw pressure measurement are not as stringent, allowing the calibration uncertainty to be as large as 2.5 Pa and the zero drift as large as 7.5 Pa [5]. In this case, at 5 m/s the yaw-pressure uncertainty normalized by the dynamic pressure is significant, $u(\Delta P_{offset}) / \Delta P_{dyn} = 0.53$.

To estimate how the uncertainty in the yaw-pressure measurement affects the yaw-null angle and the calibration factors, we conservatively specified $u(P_{offset}) = 2.5$ Pa and calculated the uncertainties $u(\theta_{NULL})$, $u(F_1)$ and $u(F_2)$ at different air speeds. The expanded uncertainty attributed *solely* to the zero-offset effect is given by

$$U(\Gamma_{q,offset}) = 2\left|\frac{\partial \Gamma_{q,FIT}}{\partial \Delta P_{23}}\right|_{NULL} u(P_{offset}) = 2\left|a_{q,1}\right| u(P_{offset}) \qquad (20)$$

where Γ_q is the generic probe variable in Table 3, $U(\Gamma_{q,offset})$ compactly represents the expanded uncertainties: $U(\theta_{NULL,offset})$, $U(F_{1,offset})$, and $U(F_{2,offset})$ for q = 1, 4, and 5, respectively. The term $[\partial \Gamma_{q,FIT}/\partial \Delta P_{23}]_{NULL}$ is the slope of the curve fit specified in Eq. (7). The slope is evaluated at the null condition where $\Delta P_{23} = 0$, and therefore equals the fit coefficient $a_{q,1}$.

Figure 10 shows the expanded uncertainties $U(\theta_{NULL,offset})$, $U(F_{1,offset})$, and $U_r(F_{2,offset})$ plotted versus the pitch angle (α) for $u(P_{offset}) = 2.5$ Pa. As denoted by the legend at the top of the figure, each symbol corresponds to a constant airspeed between 5 m/s and 30 m/s. Results for the spherical probe (left) and the prism probe (right) are plotted on the same *x* and *y* scale so that they can be easily compared. Results for the prism probe and spherical probe are nearly identical at all airspeeds, except for $U(F_{1,offset})$ at the lowest airspeed. The dependence on pitch is small, but, in general the uncertainty slightly increases at large negative or positive pitch angles. As expected, the largest uncertainties occur at the lowest airspeed of 5 m/s where the dynamic pressure is a larger fraction of $u(P_{offset})$. The uncertainty in $U(\theta_{NULL,offset})$ and $U(F_{2,offset})$ are substantial at 5 m/s, approximately 5 ° and 10 %, respectively for pitch angles ranging from $\alpha = -20$ to 20 . The uncertainties are lower at 10 m/s, but still are significant being approximately 1.5 ° and 1.5 %, respectively.

These results suggest that zero-offset effects can make a significant contribution to the uncertainty budget in field applications at airspeeds at or below 10 m/s. To improve the uncertainty requires more accurate yaw-pressure transducers and ensuring that the null parameters are measured at $\Delta P_{23} = 0$.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

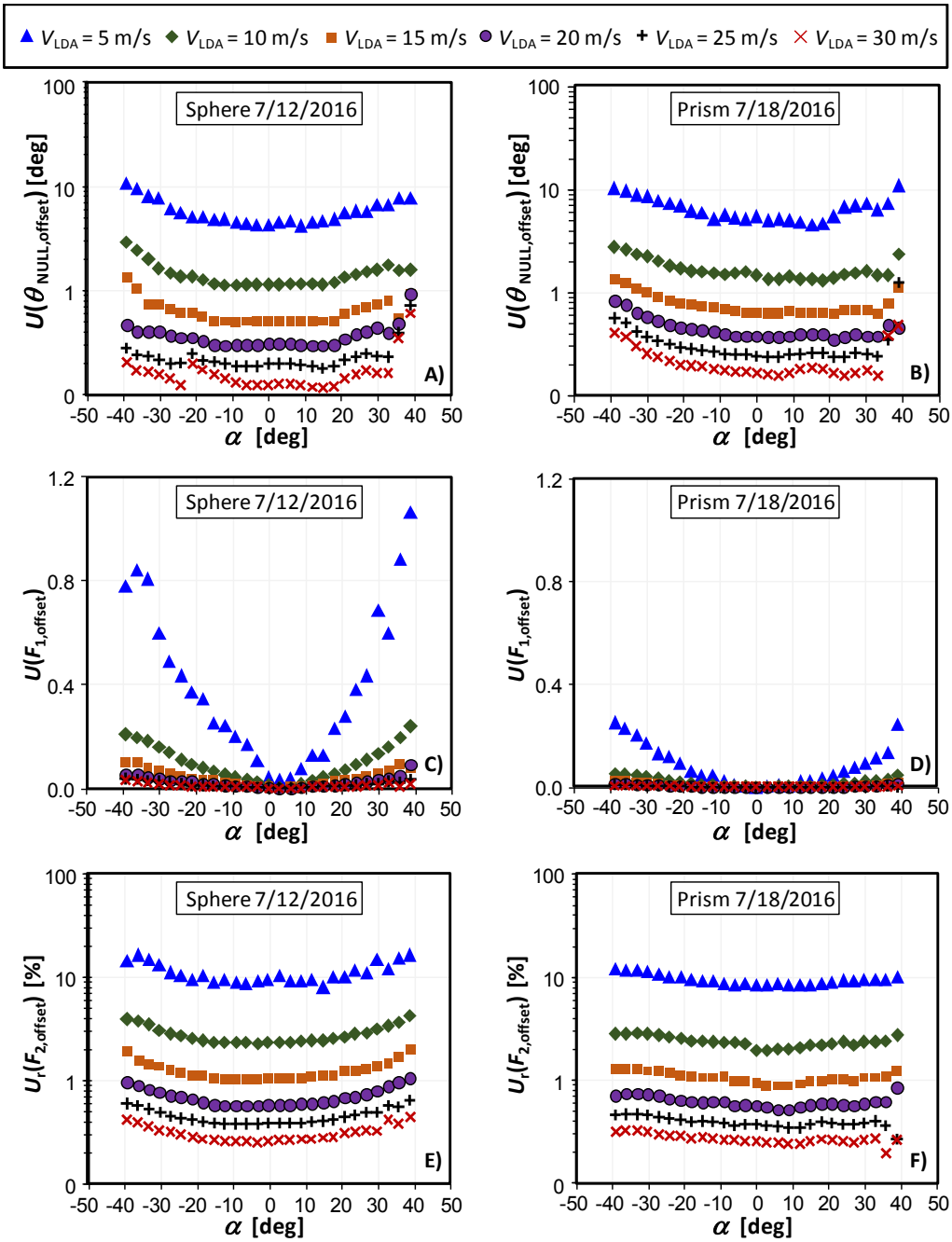


Figure 10. The expanded uncertainties in θ_{NULL}, F_1, and F_2 attributed an uncertainty in the zero-offset pressure $u(P_{offset})$ = 2.5 Pa for the spherical probe (left) and prism probe (right) for airspeeds from 5 m/s to 30 m/s and pitch angles from −40 to 40 .

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

7 DISCUSSION

We calibrated a spherical probe and prism probe using a yaw-nulling method for pitch angles $-45 \le \alpha \le 45$ and air speeds 5 m/s $\le V_{LDA} \le$ 30 m/s. Here, we focus on the narrower range $-20 \le \alpha \le 20$ which includes the conditions encountered in most industrial smokestacks. We determined the pitch calibration factor F_1 and the velocity calibration factor F_2 over the range of calibration conditions. Three repeated calibrations of the spherical probe and two repeated calibrations of the prism probe were mutually consistent within ±0.75 % of F_2. The dependence of F_2 on the Reynolds number (*Re*) differed significantly between the two probes. For the spherical probe, F_2 was nearly *Re*-independent (deviations from the mean were less than ± 1 %) except at the lowest air speed of 5 m/s, where deviations increased to nearly 2 %. In contrast, F_2 for the prism probe had a strong *Re*-dependence and the deviations from the mean were as large as 6 %. *Based on this result, the prism probe should be calibrated over the range of airspeeds for which it will be used to achieve accuracies better than 6 %*.

To apply the NIST calibration results to flue gas measurements in the field, we developed a calibration curve that determines α and F_2 from measurements of the pitch pressure and velocity pressure at the null condition. This calibration curve accounts for the *Re*-dependence of the prism probe. In contrast, EPA's protocol 2F, which is widely used for stack measurements, does not account for *Re* effects. If method 2F is implemented using the prism probe calibrated herein, velocity errors exceeding 5 % could occur. The NIST calibration curve accounts for the *Re*-dependence of the prism probe and is accurate to better than 1 % for airspeeds in the range 5 m/s $\le V \le$ 30 m/s and pitch angles from -20 to 20 .

We determined the yaw-null angle θ_{NULL} and the null pitch pressure $\Delta P_{45,NULL}$ and the null velocity pressure $\Delta P_{12,NULL}$ using a *Curve Fit Method*. We rotated the probe about its axis and measured yaw angles θ, pitch pressures ΔP_{45}, velocity pressures ΔP_{12}, and yaw pressures ΔP_{23} for values of ΔP_{23} surrounding the null condition $\Delta P_{23} = 0$. We fitted polynomial curves to three sets of data pairs: ($\Delta P_{23}, \theta$); (ΔP_{23}, ΔP_{45}); and (ΔP_{23}, ΔP_{12}). Then, we determined θ_{NULL}, $\Delta P_{45,NULL}$, and $\Delta P_{12,NULL}$ by evaluating the fitted curves at $\Delta P_{23} = 0$. This procedure avoids errors resulting from zero-offsets of pressure transducers, which often occur in field applications when the null conditions are measured at non-zero yaw pressures. To estimate the size of zero-offset effects, we multiplied the EPA's yaw pressure specification (2.5 Pa) for stack emission flow measurements by the slope of the fitted curves. The zero-offset effects increase significantly at lower airspeeds. At 5 m/s the zero-offset for the spherical probe is 10 %; for the prism probe, it ranges from 8.5 % to 10 %, depending on the pitch angle. At 10 m/s the zero-offset was 2.5 % for the spherical probe and 2 % for the prism probe.

The expanded percentage uncertainty at a 95 % confidence level of F_2 was comparable for both probes. For the prism probe, $U(F_2)$ ranged from 0.9 % to 1.25 % depending on airspeed; for the spherical probe, $U(F_2)$ ranged from 0.9 % to 1.1 %. For both probes the largest uncertainty source was the reproducibility of repeated calibrations and the LDA standard; together, they contributed 60 % to 80 % of $U(F_2)$.

8 REFERENCES

[1] Shinder, I. I., Khromchenko, V. B., and Moldover, M. R., ***NIST's New 3D Airspeed Calibration Rig, Addresses Turbulent Flow Measurement Challenges,*** 9th International Symposium on Fluid Flow Measurement (ISFFM), Washington, DC, U. S., April 2015.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

[2] Norfleet, S. K., Muzio L. J., and Martz T. D., ***An Examination of Bias in Method 2 Measurements Under Controlled Non-Axial Flow Conditions***, Report by RMB Consulting and Research, Inc., *May 22, 1998.*

[3] Borthwick, R., and Bundy, M., ***Quantification of a precision point source for generating carbon dioxide emissions***, EPRI CEM User Group Conference and Exhibit, Chicago, IL, June 8, 2011.

[4] Quick, J. C., ***Carbon dioxide emission tallies for 210 U.S. coal-fired power plants: A comparison of two accounting methods****,* Journal of the Air & Waste Management Association, January 2014.

[5] Environmental Protection Agency, 40 CFR Part 60 ***Application Method 2F Stack Gas-Velocity/Volumetric Flow Rate-3D probe***, Washington D.C., 1996.

[6] Shinder, I. I., Crowley, C. J., Filla, B. J., Moldover, M. R., ***Improvements to NIST's Air Speed Calibration Service***, Flow Measurement and Instrumentation, Vol. 44, pp. 19–26, 2015.

[7] Yeh, T. T., Hall, J. M., ***An Uncertainty Analysis of the NIST Airspeed Standards, ASME Paper FEDSM2007-37560***, ASME/JSME 5th Joint Fluids Engineering Conference, pp. 135-142, San Diego, California, USA, 2007.

[8] Yeh, T. T. and Hall, J. M. ***Uncertainty of NIST Airspeed Calibrations***, *Technical document of Fluid Flow Group*, NIST, Gaithersburg, Maryland, USA, 2008.

[9] Yeh, T. T., Hall, J. M., ***Air Speed Calibration Service*****, NIST Special Publication 250-79**, National Institute of Standards and Technology, Gaithersburg, Maryland, 2006.

[10] Shinder, I. I., Crowley, C. J., Filla, B. J., Moldover, M. R., ***Improvements to NIST's Air Speed Calibration Service***, 16th International Flow Measurement Conference, Flomeko 2013, Paris, France September 24-26, 2013.

[11] Jaeger, K. B. and Davis, R. S. ***A Primer for Mass Metrology***, National Bureau of Standards (U.S.), Special Publication 700-1, Industrial Measurement Series. 1987: 79.

[12] Taylor, B. N. and Kuyatt, C. E., ***Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results***, NIST Technical Note 1297, National Institute of Standards and Technology, January 1993.

[13] International Organization for Standardization, ***Guide to the Expression of Uncertainty in Measurement***, Switzerland, 1996 edition.

[14] Coleman, H. W. and Steele, W. G., ***Experimentation and Uncertainty Analysis for Engineers***, 3rd ed., John Wiley and Sons, Inc., N.Y., 2009.

[15] DuChateau, P. and Zachmann, D., ***Applied Partial Differential Equations***, Harper and Row Publishers Inc., New York, N.Y., 1989.

[16] Kennedy, John B. and Adam M. Neville, ***Basic Statistical Methods for Engineers and Scientists,*** 3rd ed., Harper and Row, 1986.

Filla, Bernard; Johnson, Aaron; Khromchenko, Vladimir; Moldover, Michael; Shinder, Iosif.
"Characterization of Five-Hole Probes used for Flow Measurement in Stack Emission Testing."
Paper presented at 2018 International Symposium on Fluid Flow Measurement (ISFFM), Querétaro, Mexico. March 21, 2018 - March 23, 2018.

Through-focus Scanning Optical Microscopy Applications

Ravi Kiran Attota*
Engineering Physics Division, National Institute of Standards and Technology
Gaithersburg, MD 20899, USA

ABSTRACT

We present a partial application space of the metrology method referred to as through-focus scanning optical microscopy (TSOM), with most number of favorable attributes as a metrology and process control tool. TSOM is a NIST-developed, high-throughput and low-cost optical metrology tool for dimensional characterization with sub-nanometer measurement resolution of nano-scale to microscale targets using conventional optical microscopes, with many unique benefits and advantages. In TSOM the complete set of out-of-focus images are acquired using a conventional optical microscope and used for dimensional analysis. One of the unique characteristics of the TSOM method is its ability to reduce or eliminate optical cross correlations, often challenging for optical based metrology tools. TSOM usually has the ability to separate different dimensional differences (i.e., the ability to distinguish, for example, linewidth difference from line height difference) and hence it is expected to reduce measurement uncertainty.

TSOM is applicable to a wide variety of target materials ranging from transparent to opaque, and shapes ranging from simple nanoparticles to complex semiconductor memory structures, including buried structures under transparent films. TSOM has been successfully applied to targets ranging from one nm to over 100 μm (over five orders or magnitude size range). Demonstrated applications of TSOM include critical dimension (linewidth), overlay, patterned defect detection and analysis, FinFETs, nanoparticles, photo-mask linewidth, thin-film (less than 0.5 nm to 10 nm) thickness, through-silicon vias (TSVs), high-aspect-ratio (HAR) targets and others with several potential three-dimensional shape process monitoring applications such as MEMS/NEMS devices, micro/nanofluidic channels, flexible electronics, self-assembled nanostructures, and waveguides. Numerous industries could benefit from the TSOM method —such as the semiconductor industry, MEMS, NEMS, biotechnology, nanomanufacturing, nanometrology, data storage, and photonics.

Keywords: TSOM, through focus, optical microscope, nanometrology, process control, nanomanufacturing, nanoparticles, overlay metrology, critical dimension, defect analysis, dimensional analysis, MEMS, NEMS, photonics

1. INTRODUCTION

As the applications of nanotechnology and micro-technology become widespread, three-dimensional (3-D) shape analysis of microscale to nanoscale structures with sub-nanometer measurement resolution become important. Several tools and methods are available for such a metrology with their unique advantages and disadvantages. However, it is a challenge to find a universally applicable measurement tool. But an economical, versatile and high-throughput tool that provides 3-D shape measurement with least number of negatives is highly desirable, especially for industrial nanomanufacturing. Here we present some of the applications of the metrology method referred to as through-focus scanning optical microscopy (TSOM), with most number of favorable attributes as a metrology and process control tool [1-27].

TSOM is a NIST-developed, high-throughput and low-cost optical metrology tool for dimensional characterization with sub-nanometer measurement resolution of nano-scale to microscale targets using conventional optical microscopes, with many unique benefits and advantages. For this reason TSOM has been well recognized: given an R&D 100 award in 2010 [28], included in the International Technology Roadmap for Semiconductors (ITRS) [29, 30] and International Roadmap for Semiconductor Devices and Systems (IRDS) 2017 Edition: Metrology [31], for several applications, included in the SEMI document for 3-D S-IC structures [32], identified by SEMATECH for several applications[33].

*Ravikiran.attota@nist.gov; Phone: 1 301 975 5750

Attota, Ravikiran.
"Through-focus Scanning Optical Microscopy Applications."
Paper presented at SPIE Photonics Europe, Strasbourg, France. April 22, 2018 - April 26, 2018.

In conventional optical microscopy, out-of-focus images are ordinarily considered not particularly useful, especially for metrology applications. This is based on the generally accepted assumption that optical information from out of focus planes is either less useful or detrimental compared to information from the best focus plane. However, a complete set of out-of-focus images contains significantly more information about the target as compared to a single best-focus image. This additional information can be extracted by TSOM. A typical TSOM image is a cross-section constructed from the four-dimensional (4-D) optical data acquired using a conventional optical microscope as a target is scanned along the focus direction.

TSOM shows considerable promise in meeting the industrial requirement of high throughput and low-cost 3-D shape metrology. TSOM transforms a conventional and ubiquitous optical microscope into a powerful 3-D shape metrology tool and provides sub-nanoscale measurement resolution comparable to SEM and AFM in both lateral and vertical directions. One of the unique characteristics of the TSOM method is its ability to reduce or eliminate optical cross correlations, often challenging for optical based metrology tools. TSOM usually has the ability to separate different dimensional differences (i.e., the ability to distinguish, for example, linewidth difference from line height difference) and hence it is expected to reduce measurement uncertainty.

TSOM is applicable to a wide variety of target materials ranging from transparent to opaque, and shapes ranging from simple nanoparticles to complex semiconductor memory structures, including buried structures under transparent films.

TSOM has been successfully applied to targets ranging from one nm to over 100 μm (over five orders or magnitude size range). Demonstrated applications of TSOM include critical dimension (linewidth), overlay, patterned defect detection and analysis, FinFETs, nanoparticles, photo-mask linewidth, thin-film (less than 0.5 nm to 10 nm) thickness, through-silicon vias (TSVs), high-aspect-ratio (HAR) targets and others with several potential three-dimensional shape process monitoring applications such as MEMS/NEMS devices, micro/nanofluidic channels, flexible electronics, self-assembled nanostructures, and waveguides. Numerous industries could benefit from the TSOM method —such as the semiconductor industry, MEMS, NEMS, biotechnology, nanomanufacturing, nanometrology, data storage, and photonics. Here we highlight some examples for which TSOM was applied.

2. TSOM IMAGING METHOD

A brief introduction to the creation of TSOM image from the through-focus optical data is presented here. First, a set of through-focus images are collected by scanning the stage on which a target is placed along the axial direction (Fig. 1, left image). Stacking the set of through-focus images at their respective focus positions creates a 3-D space filled with the optical intensity data (Fig. 1, center image). A vertical cross-section through this 3-D space creates a TSOM image (Fig. 1, right image). In the TSOM image, the X and Y axes represent distance, and focus position, respectively.

A differential TSOM (D-TSOM) image is a pixel-by-pixel difference between two normalized and aligned TSOM images [12]. If the two TSOM images are obtained using two targets with slightly different dimensions, then the D-TSOM image color pattern highlights the type of dimensional difference between the two targets [3, 9, 12]. If the two TSOM images are obtained using the same target (similar to repeated collection), then the D-TSOM image highlights noise [9, 15, 16], including optical and vibrational noise.

One way to quantify optical content of a D-TSOM (or a TSOM) image is with an optical intensity range (OIR) metric which is defined as the absolute difference between the maximum and the minimum optical intensity in a given D-TSOM (or a normalized TSOM) image and multiplied by 100 [12]. OIR represents the magnitude of the optical intensity difference. A second way is by evaluating the mean absolute value (MAV), calculated as:

$$MAV = \frac{1}{n}\sum_{k=1}^{n} |D\text{-}TSOM|_k$$

where N is the number of pixels in the D-TSOM image. MAV represents total optical content in a D-TSOM (or a TSOM) image. In the current work all MAVs presented are multiplied by 1000.

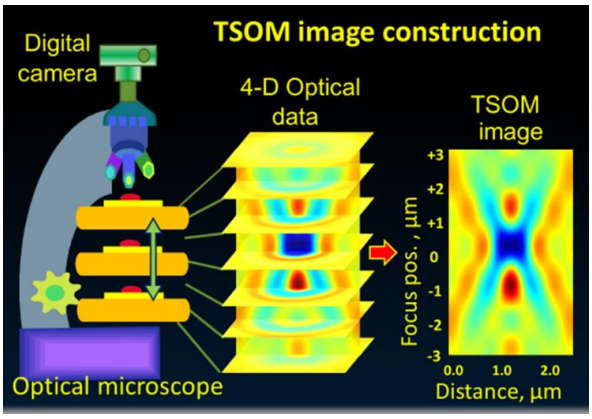


Figure 1. Video depicting a typical TSOM image construction process. http://dx.doi.org/doi.number.goes.here

3. APPLICATIONS OF TSOM

3.1 Defect inspection of EUV masks

The defect density of extreme ultraviolet (EUV) masks has been one of the top concerns in EUV lithography[10]. To reduce the defect level of EUV masks, defect sources must be mitigated or eliminated at each step of the mask manufacturing process. Defects on EUV masks are categorized as phase defects and amplitude defects based on the phase change or amplitude change of reflected EUV light. Amplitude defects are located on top or near the top of the multilayer structure. Phase defects are generally formed by the multilayer deposition over substrate defects (particles or pits) or are particles added during multilayer deposition.

TSOM demonstrated potential capability of detecting defects on EUV masks down to 15 nm size, and it may be able to determine depth location, with information on amplitude and phase change of light. Figure 2(a) is a simulation, using optical imaging modeling software, of applying the TSOM method to detect a bump-type phase defect having 21 nm spherical-equivalent volume diameter (SEVD). Simulations indicate that the TSOM method should have sufficient signal strength over experimental noise to detect phase defects down to 15 nm SEVD using 193 nm wavelength light. A pit type of defect exhibits distinct color reversal in comparison to the bump type of defect as shown in Fig. 2(b), clearly distinguishing the two.

Nondestructively detecting the depth of buried defects (pits or particles) in a multilayer EUV stack is a challenging task, but the simulated TSOM images for such buried particle defects do show sensitivity to the depth of particles. This suggests the possibility of inferring the depth of defects in multilayers, which is essential information for defect repair schemes. We are in the process of experimentally validating the simulation results by building a controlled blank EUV mask test structure having particles of known size placed at different known depths in the multilayer structure. This validation will be extremely beneficial for the defect repair process, resulting in major savings in time and cost. TSOM shows potential to (i) detect phase defects on EUV masks down to 15 nm SEVD, (ii) distinguish pits from bumps, and (iii) infer depth information of buried particles. The TSOM method is economical as it requires a low cost 193 nm optical microscope (compared to an AIT), and has high throughput. As a result, the TSOM method appears to be a good candidate method for detection of defects on EUV lithography masks. For more information refer to [10].

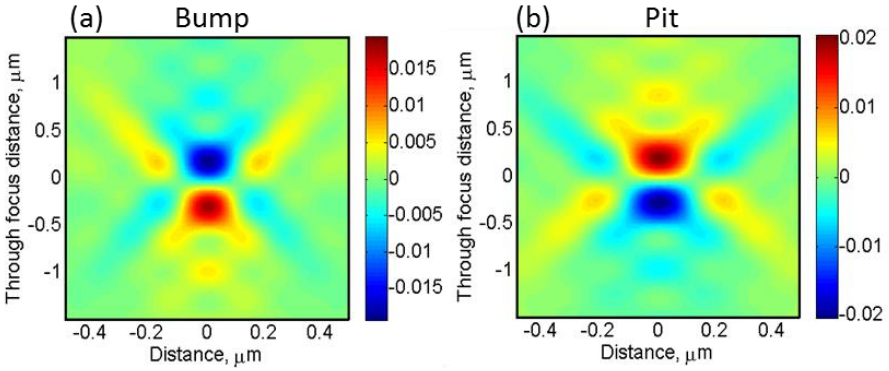


Figure 2. Simulated TSOM images of (a) bump type and (b) pit type 21 nm SEVD phase defects showing color reversal. λ = 193 nm.

3.2 Nanoparticle size determination

There is a great interest in using nanoparticles [13, 34, 35] because of their variety of biological applications. Size of the nanoparticles is a critical factor for their effective biological application. Although there are several tools available for size determination of nanoparticles, there is a concern that most of the available high-throughput methods have insufficient level of accuracy as per the definition of particle size distribution by the National Institute of Standards and Technology. In addition, there is also a concern that "soft" or "fuzzy" nanoparticles are difficult to characterize using high-resolution methods such as transmission electron microscopy (TEM) or scanning electron microscopy (SEM), since these methods deform nanoparticles during the measurement process. The most widely used dynamic light scattering (DLS) method also has several limitations such as systematic offset compared with TEM measurements.

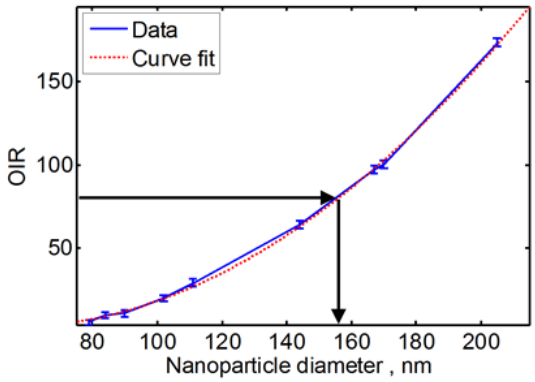


Figure 3. Calibration curve generated using the measured TSOM images of spherical Au nanoparticles on a Si substrate. Arrow marks indicate size determination of an "unknown" spherical Au nanoparticle using the OIR value of the TSOM image. Wavelength = 520 nm, σ = 0.176 (Illumination NA = 0.15, collection NA = 0.85).

Applicability of TSOM was tested for spherical Au nanoparticles. Using an SEM, spherical-shaped Au nanoparticles spread on a Si substrate were identified and their sizes were measured. The identified Au nanoparticles were then analyzed using the TSOM method. OIR values from the TSOM images were plotted as a function of the SEM-measured mean particle diameters. A fitted curve using these values results in a calibration curve (or library) as shown in Fig. 3 for the Au nanoparticles under the experimental conditions used. The 156 nm diameter nanoparticle was considered as a test target and hence its value was not included in the calibration curve. The size of the test target was measured by comparing its OIR value with the calibration curve. As shown in Fig. 3, the TSOM-measured diameter is 153 nm, while the SEM-measured diameter is 156 nm. For more information refer to [13].

3.3 Determine the number of nanoparticles in a cluster

There are a few optical microscopy techniques to identify the number of the nanoparticles by measuring the optical

properties of the nanoparticles such as the fluorescence lifetime, spectrum, and time-dependent intermittence [36]. However, all these optical microscopy techniques are restricted to the fluorescent nanomaterials. High-resolution imaging tools such as scanning electron microscope (SEM), transmission electron microscope, and atomic force microscope are capable of counting the number of the nanoparticles in a cluster. However, they do not have the functionality to measure photophysical properties. Most photophysical properties of the nanoparticles are measured with optical microscopes [37]. Therefore, adding the functionality to determine the number of nanoparticles by the optical microscope gives a significant advantage to users and also helps them quantitatively interpret the photophysical properties of the nanoparticle clusters.

After spreading 93 nm mean size polystyrene nanoparticles onto a Si substrate, monomers, dimers, trimers, and tetramers were identified using SEM. The same clusters were analyzed using TSOM and their OIR values were plotted as shown in Fig. 4. We can observe that the optical content (OIR) of the TSOM images increased with the number of the particles in the clusters. To confirm the validity of the trends observed from the measurements, we also simulated optical TSOM images under the exact experimental conditions. These results are also plotted in Fig. 4. Even though there are some differences between the measurements and the simulations, the overall trend matches very well as a function of the number of particles in the clusters and thus provides confidence in the process. For more information refer to [14].

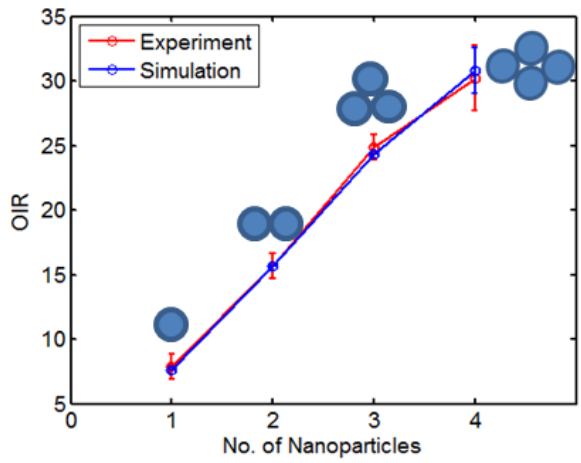


Figure 4. A plot of the OIR values from the measurements and the simulations for the different clusters of polystyrene nanoparticles

3.4 Resolving three-dimensional shape of sub-50 nm wide lines

The ability to meaningfully resolve (or distinguish) three-dimensional (3-D) shape and variations at the nanometer scale is extremely beneficial in the current world of nanotechnology. There is a significant need for tools with the ability to resolve objects at the nanometer scale while also being economical, versatile, robust, easy-to-use, non-destructive, non-contaminating, and non-contact, yet capable of high throughput. It is additionally beneficial if the tool has the ability to resolve feature shape in 3-D. We experimentally demonstrate that the three-dimensional (3-D) shape variations of nanometer-scale objects can be resolved and measured with sub-nanometer scale sensitivity using conventional optical microscopes by analyzing 3-D optical data using TSOM.

The results show that TSOM-determined cross-sectional (3-D) shape differences of 30 nm to 40 nm wide lines agree well with critical-dimension atomic force microscope measurements (Fig. 5). The TSOM method showed a linewidth uncertainty of 1.22 nm (k=2). Complex optical simulations are not needed for analysis using the TSOM method, making the process simple. For more information refer to [12].

Simulations demonstrate the potential usefulness of TSOM for 3-D shape metrology of fin structures for the 22 nm and 16 nm nodes. TSOM shows the ability to detect sub-nanometer, three-dimensional shape variations such as line height, sidewall angle, width, and pitch in fins of fin-shaped field effect transistor structures using conventional optical microscopes. Sensitivity to various perturbations appears sufficient and distinct (Table. 1). In addition to economical metrology hardware, the substantially smaller area targets needed result in further potential cost savings. If paired with CD-SEM, scatterometry (optical critical dimension) or other widely-used tools, an inexpensive tool such as TSOM could be a key component of a good hybrid metrology solution. For more information refer to [9].

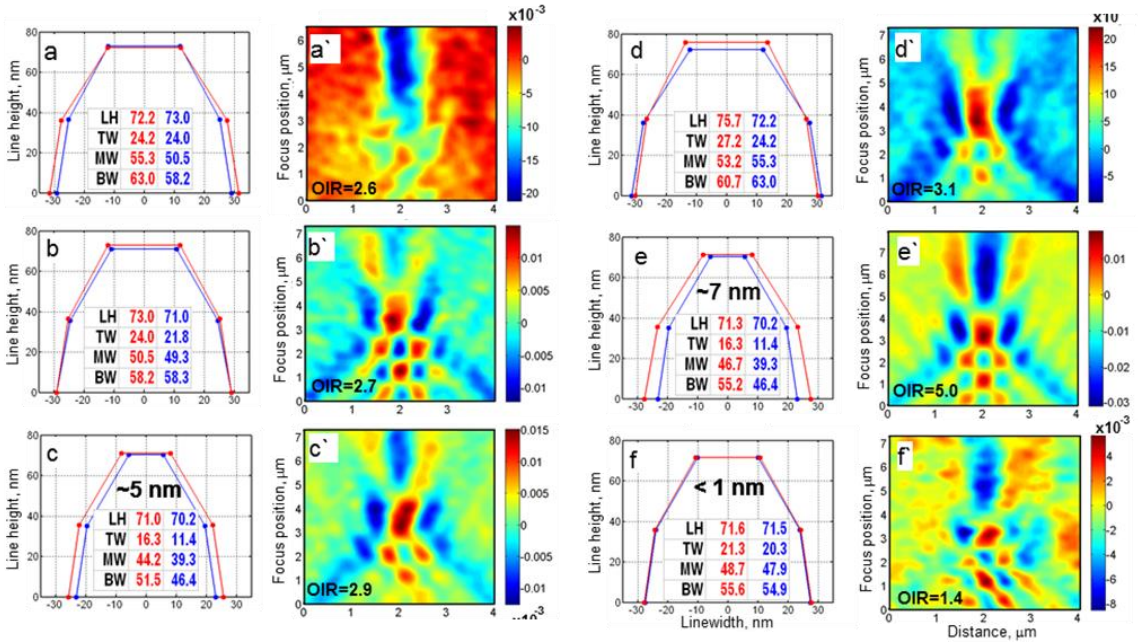


Figure 5. Experimental results for size and shape analysis of isolated lines. (a,b,c,d,e,f) Plots showing the CD-AFM-measured dimensions of the pair of lines selected for TSOM analysis. Solid dots indicate the locations of the CD-AFM measured values. Line height (LH), top width (TW), middle width (MW) and bottom width (BW, in nanometers) are shown in order for the pair of lines. (a`,b`,c`,d`,e`,f`) are the corresponding DTIs along with their OIR values for the selected pairs of lines. Different types of shape changes (a,b,c,d,) produced DTIs with distinctly different patterns (a`,b`,c`,d`). Varying magnitudes of the same shape difference as indicated in c,e,f produced similarly patterned DTIs (c`,e`,f`). The OIR of the DTI signal, shown by the color bar scale, is greater for larger differences in the linewidth values.

ITRS Node	λ	LW (CD)	SWA	Box recess	LH	PW	High K thickness
nm	nm	nm	Degree	nm	nm	nm	nm
22	546	**0.04**	**0.06**	**0.28**	0.08	**0.44**	?
22	248	0.21	0.38	>3	0.15	1.11	?
22	193	**0.08**	**0.1**	>2	**0.04**	**0.2**	**0.05**
16	546	**0.07**	0.26	>3	0.24	**0.14**	?
16	258	**0.12**	0.45	>3	**0.2**	**0.41**	?
16	193	**0.13**	**0.27**	>3	**0.15**	0.89	?

Table 1. Summary of results of varying features at a given ITRS node and TSOM wavelength. The value of the smallest perturbation that can be detected over noise is given in the table. Underline = no sensitivity, Italics = a high likelihood of cross-correlation; Normal = better sensitivity, and Bold = the best sensitivity.

3.5 Optical microscope illumination analysis

Illumination distortion in an optical microscope can lead to quantitative misinterpretation of measurements that are based on through-focus images, such as confocal microscopy [38], 3-D super-resolution imaging of nanoparticles [39], optical nanoparticle and bacteria tracking methods [40-44], and the TSOM method[2, 21, 27]. Similar to TSOM, the optical tracking methods also use a low illumination NA [39, 44], making the current study even more relevant to them also. We speculate that the slanted TSOM image reported in [21] could be due to distorted illumination, probably caused by illumination aperture misalignment. Hence, it is advisable to measure and correct the distortions in the illumination to make a proper analysis for these methods.

Distortions in illumination can be investigated by evaluating the angular illumination asymmetry (ANILAS) maps [45, 46]. Here we show that such aberrations in the illumination also produce distorted TSOM images, namely lateral image shift with focus height. We demonstrate here that the slant in TSOM images caused by illumination aberrations of this type can nearly be eliminated by simply re-aligning the aperture diaphragm to the optical axis. Conversely, the slanted TSOM images could provide an alternative, more convenient way of measuring ANILAS.

For example, illumination distortions at the sample plane of an optical microscope can be artificially created by moving the aperture diaphragm along the optical axis as shown in Figs. 6(a1) and 6(b1). This results in illumination distortions as shown by ANILAS maps in Figs. 6(a2) and 6(b2). There is a good correlation between the ANILAS map and the slant in the TSOM image. The TSOM image at the center of the ANILAS map (lowest asymmetry) is upright (Fig. 6(ab4)), indicating that no matter where the "sweet spot" is located in the FOV, the through-focus data at that location provides a more undistorted result. This also shows the importance of determining the location of the "sweet spot" in the FOV. For such an asymmetric illumination condition, the TSOM image of a vertical line is slanted despite being at the center of the FOV (Fig. 6(b4)), similar to the illumination condition in Fig. 6(a4). However, the type of slant in the TSOM image (either left or right) depends on the location of the line target with respect to the center ANILAS region. The slant in the TSOM images are toward the left and the right, for the line targets located on the left and on the right sides of the center of the lowest-ANILAS region, respectively. Hence by observing and analyzing the slant in a TSOM image, illumination distortions can be determined. For more information refer to [19].

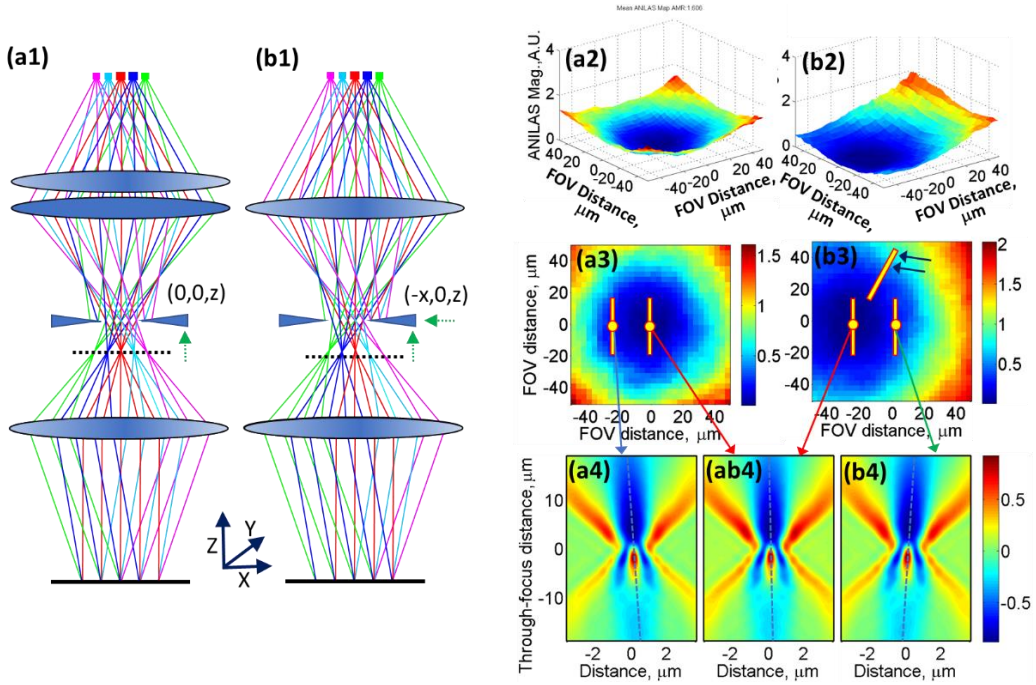


Figure 6. Schematics showing angular illuminations at the sample plane with the aperture diaphragm moved along the optical axis toward the field diaphragm for (a1) no lateral shift (X = 0, Y = 0, Z = z), and (b1) with finite lateral (left) shift in the X direction (X = - x, Y = 0, Z = z). (a2) (a3) and (a4) are the 3-D ANILAS map, 2D ANILAS map and TSOM image, respectively, for the illumination condition shown in (a1) with no lateral shift (X = 0, Y = 0, Z = (1250 ± 5) μm). (a4) was obtained by placing the line at the location and in the orientation as shown by the arrow mark in (a3). (b2), (b3) and (b4) are the 3-D ANILAS map, 2D ANILAS map and TSOM image, respectively, for the illumination condition shown in (b1) with (20 ± 2) μm lateral left shift (X = (-20 ± 2 μm), Y = 0, Z = (1250 ± 5) μm). (b4) was obtained by placing the line vertically at the location as shown by the arrow mark in (b3). (ab4) represents a typical upright TSOM image obtained by placing the vertically oriented line at the locations as indicated in (a3) and (b3) by red arrow marks. All the TSOM images were obtained using a 950 nm wide line.

3.6 Overlay Analysis

An overlay target was designed for double patterning. Using both simulations and experimental measurements, the TSOM method demonstrated the potential for near 0.1-nm measurement resolution. The design consists of line gratings with alternate lines formed during two process steps. The complete overlay target consists of at least five adjacent gratings with five designed overlay offsets as shown in Fig. 7(a). A simulated TSOM image of such an overlay target is shown in Fig. 7(b) for zero process overlay. Optical contrast of the TSOM image increases with a greater overlay offset. The TSOM image changes as shown in Fig. 7(c) for a 2 nm process overlay offset. Optical intensities can be integrated in each designed overlay section of the target and plotted as a function of the designed overlay offset as shown in Fig. 7(d). In these plots, the minimum point indicates the process overlay offset.

Overlay targets were fabricated based on the composite overlay design. A typical optical image of one such experimental composite overlay target is shown in Fig. 7(e) for zero process overlay, showing very little signal. The measured TSOM image using such a target is shown in Fig. 7(f). As expected, optical contrast increases with greater overlay and validates the simulation results. The composite overlay targets are self-calibrating, robust to process variations and optical

aberrations, and highly sensitive, potentially providing near 0.1-nm overlay measurement resolution for double or multiple patterned process applications. TSOM-based analysis for other types of overlay targets can be found in Ref. [47].

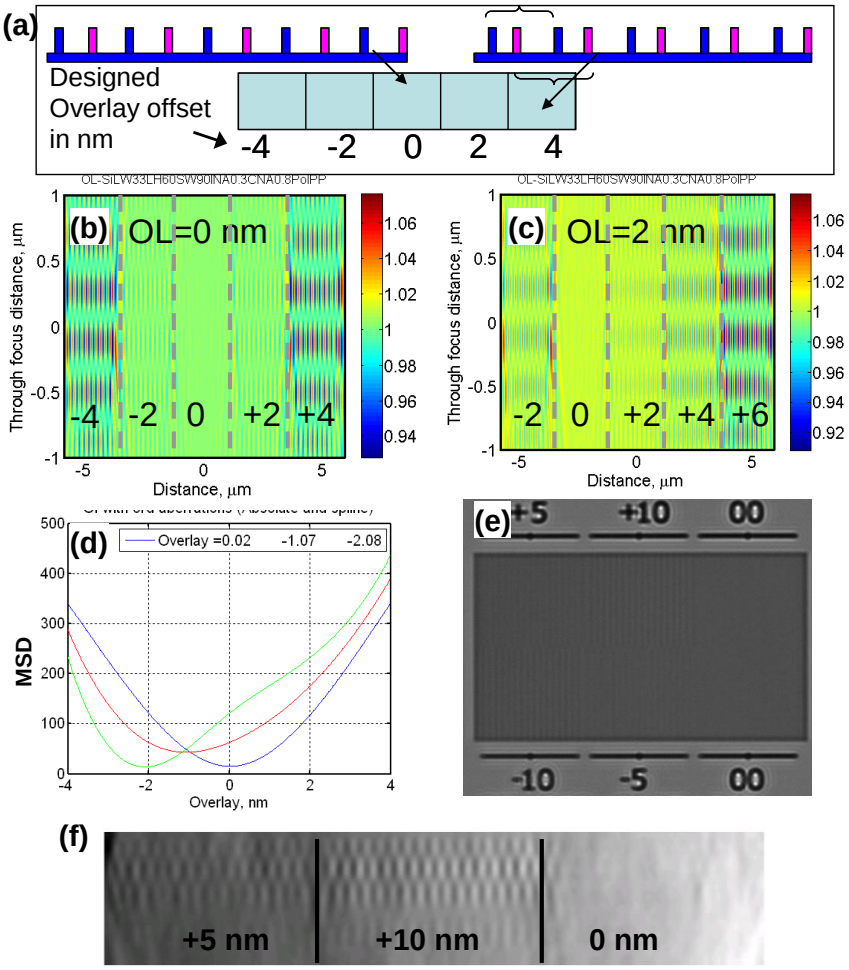


Figure 7. (a) Composite overlay target design composed of multiple designed overlay offsets as indicated for double patterning. Simulated TSOM images of the composite overlay target for (b) zero and (c) 2 nanometer overlay offsets showing increased image contrast with increased overlay offsets (λ = 193 nm) . (d) A plot of integrated optical intensities (MSD) as a function of overlay offset (from simulations). Experimentally acquired (e) optical image and (f) TSOM image of a composite overlay target using λ = 546 nm (designed overlay offsets shown are in nm).

3.7 Dimensional Analysis of Through-Silicon Vias (TSVs)

TSVs require truly 3-D metrology tools for dimensional analysis as they extend as deep as 50 μm with diameters as large as 5 μm. The TSOM method is ideally suited for dimensional analysis of large targets such as TSVs. Simulated differential TSOM images for diameter and depth change are shown in Fig. 8. As expected, a change in diameter (Fig. 8(a)) results in a distinctly different signal than a change in depth (Fig. 8(b)). Based on the signal strength in these simulated differential

images, it may be possible to detect changes as small as a few nanometers in the diameter and depth of 5 μm diameter and 25 μm deep TSVs. Similar type of defect analysis could be applied to MEMS devices.

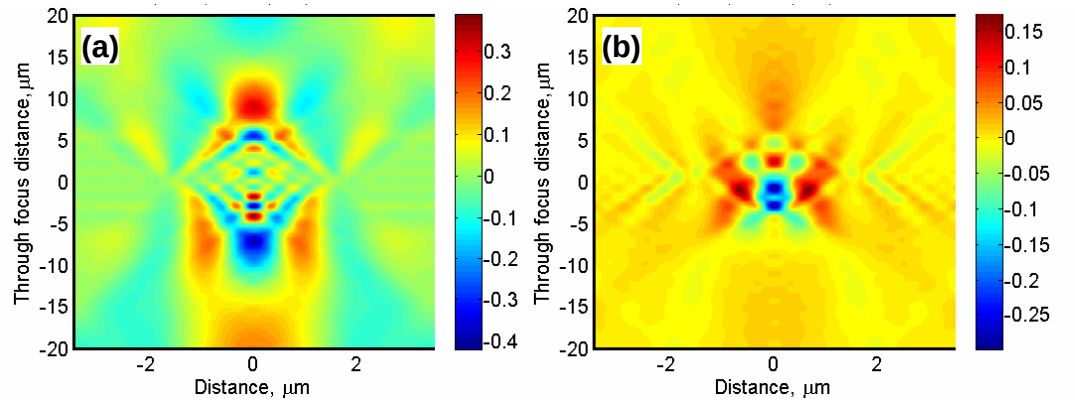


Figure 8. Simulated differential TSOM images for through-silicon vias (TSVs) (a) for 20 nm difference in the diameter, and (b) 20 nm difference in the depth of 5 μm diameter and 25 μm deep through-silicon via using λ = 546 nm.

3.8 Patterned Defect Analysis

The main challenge facing defect inspection is the expected loss of sensitivity in conventional bright field technologies. Historically, gradual improvements to bright field have successfully extended its use to smaller nodes; however, recent theoretical work indicates that this may no longer be a useful path near the 11 nm node due to significant loss of defect signal. It is therefore necessary to explore alternative approaches that may bridge this gap from the perspective of both sensitivity and throughput. To assess the defect inspection capabilities of TSOM for sub-22 nm nodes, we first constructed a simplified gate-level layout representing conventional planar gate structures with CDs and pitches of varying dimensions. For simplicity, both the structures and the underlying substrate were modeled as Si. The simulation methodology consisted of generating TSOM images for reference targets and defect targets and subsequently obtaining a differential image by subtracting intensities pixel by pixel for each condition studied. This methodology is analogous to what is done in conventional defect inspection and allows isolating the defect signal from the background. Figure 9 and table 2 depict the simulation parameters and defect structures used in this study.

Optical Parameters	
Wavelength (nm)	546, 248
Polarization angle (degrees)	0, 90
Structural Parameters	
Gate CDs (nm)	11, 15, 20
Defect Size (nm)	5 – 15
Defect Height (nm)	8 (25%), 20 (50%), 40 (100%)
Gate Pitch (nm)	60, 90, 130

Table 2. Simulation conditions for the defect structures

Although several results are available [6] here we present an important aspect of defect inspection showing how pattern density affects defect detectability. To evaluate the effects of this parameter on defect signal intensity, we performed simulations at two different pitches, 90 nm and 60 nm, for an 11 nm gate layout. All defects were simulated at three different sizes of nominal CD: 150 %, 100 %, and 50 %. Optical illumination was simulated at 248 nm and two different polarizations, 0 and 90 degrees. Figure 10 shows an excerpt of the results from this set of simulations. Once again, the modeling shows that TSOM should be capable of detecting all defects down to the smallest size (5.5 nm). In this case, all successful detection events occurred under 90 degree polarization illumination. The smallest OIR is greater than 2, showing that this optical setup might be able to detect smaller defects, perhaps at tighter pitches. The results at 0 degree polarization show once more that the type A defect produces the weakest signal, which is below the detectable limit of 1 % in at least one case for all pitch values studied.

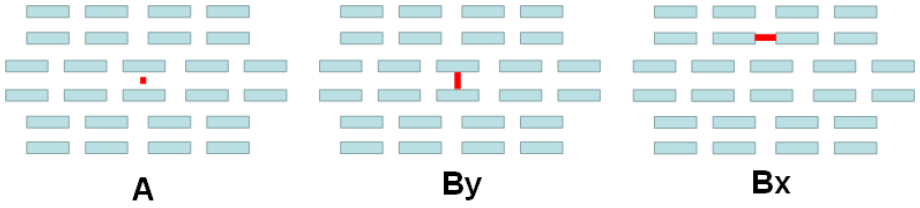
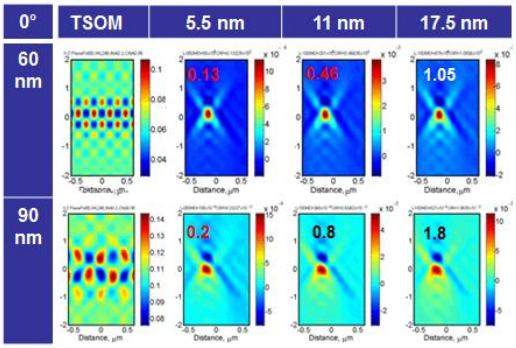
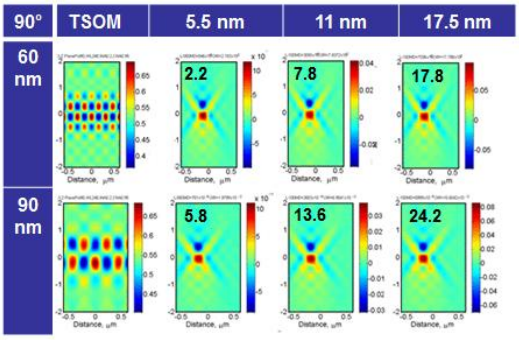


Figure 9. Summary of optical and structural parameters varied and diagrams showing the type of defects studied.





Structural Parameters	
Gate CD (nm)	11
Defect Size (nm)	5.5, 11, 17.5
Gate Pitch (nm)	60, 90
Optical Parameters	
Wavelength (nm)	248
Polarization angle (degrees)	0, 90

Figure 10. TSOM and differential TSOM images showing the effect of gate pitch on defect detectability. Images correspond to three type A defect sizes. Illumination wavelength is 248 nm, and both polarization conditions are shown. Inset numbers correspond to the OIR of the differential TSOM image. For an OIR > 1, the modeling predicts that the defect should be detectable. The table shows the set of varied parameters.

4. CONCLUSIONS

TSOM is a high-throughput and low-cost optical metrology tool for 3-D shape process monitoring and characterization. We have present several recent applications of TSOM. In addition to these we are in the process of evaluating TSOM for several other applications in diverse areas. TSOM achieved sub-nanometer measurement resolution with most number of favorable attributes as a metrology and process control tool. TSOM can be applied to sub-10 nm to over 100 μm size targets with many unique benefits and advantages. TSOM usually has the ability to separate different dimensional differences (i.e., the ability to distinguish, for example, linewidth difference from line height difference) and hence it is expected to reduce measurement uncertainty. Numerous industries could benefit from TSOM—such as the semiconductor industry, MEMS, NEMS, biotechnology, nanomanufacturing, nanometrology, data storage, and photonics.

ACKNOWLEDGEMENTS

The author would like to acknowledge the collaboration of John Kramar, Ronald Dixson, Emil Agocs, Haesung Park, Peter

Weck, Benjamin Bunday, Victor Vartanian, Heyonggon Kang, Vipin Tondare, Prem Kavuri, Andras Vladar, Rick Silver, Richard Kasica, Richard Allen, George Orji, Abraham Arceo, Vibhu Jindal, James Potzick, Michael Stocker and Lei Chen.

REFERENCES

[1] R. Attota, R. M. Silver, and J. Potzick, "Optical illumination and critical dimension analysis using the through-focus focus metric method - art. no. 62890Q," Novel Optical Systems Design and Optimization IX, 6289, Q2890-Q2890 (2006).

[2] R. Attota, T. A. Germer, and R. M. Silver, "Through-focus scanning-optical-microscope imaging method for nanoscale dimensional analysis," Optics Letters, 33(17), 1990-1992 (2008).

[3] R. Attota, R. G. Dixson, J. A. Kramar *et al.*, "TSOM Method for Semiconductor Metrology," Proc. SPIE, 7971, 79710T (2011).

[4] R. Attota, R. G. Dixson, and A. E. Vladar, "Through-focus Scanning Optical Microscopy," Scanning Microscopies 2011: Advanced Microscopy Technologies for Defense, Homeland Security, Forensic, Life, Environmental, and Industrial Sciences, 8036, (2011).

[5] R. Attota, and R. Silver, "Nanometrology using a through-focus scanning optical microscopy method," Measurement Science & Technology, 22(2), (2011).

[6] A. Arceo, B. Bunday, V. Vartanian *et al.*, "Patterned Defect & CD Metrology by TSOM Beyond the 22 nm Node," Metrology, Inspection, and Process Control for Microlithography Xxvi, Pts 1 and 2, 8324, (2012).

[7] B. Damazo, R. Attota, P. Kavuri *et al.*, "Nanoparticle Size and Shape Evaluation Using the TSOM Method," Metrology, Inspection, and Process Control for Microlithography Xxvi, Pts 1 and 2, 8324, (2012).

[8] A. Arceo, B. Bunday, and R. Attota, "Use of TSOM for sub-11 nm node pattern defect detection and HAR features," Metrology, Inspection, and Process Control for Microlithography Xxvii, 8681, (2013).

[9] R. Attota, B. Bunday, and V. Vartanian, "Critical dimension metrology by through-focus scanning optical microscopy beyond the 22 nm node," Applied Physics Letters, 102(22), (2013).

[10] R. Attota, and V. Jindal, [Inspecting mask defects with through-focus scanning optical microscopy] SPIE Newsroom, August 8 (2013).

[11] V. Vartanian, R. Attota, H. Park *et al.*, "TSV reveal height and dimension metrology by the TSOM method," Metrology, Inspection, and Process Control for Microlithography Xxvii, 8681, (2013).

[12] R. Attota, and R. G. Dixson, "Resolving three-dimensional shape of sub-50 nm wide lines with nanometer-scale sensitivity using conventional optical microscopes," Applied Physics Letters, 105(4), (2014).

[13] R. Attota, P. P. Kavuri, H. Kang *et al.*, "Nanoparticle size determination using optical microscopes," Applied Physics Letters, 105(16), (2014).

[14] H. Kang, R. Attota, V. Tondare *et al.*, "A method to determine the number of nanoparticles in a cluster using conventional optical microscopes," Applied Physics Letters, 107(10), (2015).

[15] R. Attota, "Noise analysis for through-focus scanning optical microscopy," Optics Letters, 41(4), 745-748 (2016).

[16] R. Attota, and J. Kramar, "Optimizing noise for defect analysis with through-focus scanning optical microscopy," Proc. SPIE, 9778, 977811 (2016).

[17] R. K. Attota, and H. Kang, "Parameter optimization for through-focus scanning optical microscopy," Optics Express, 24(13), 14915-14924 (2016).

[18] R. K. Attota, P. Weck, J. A. Kramar *et al.*, "Feasibility study on 3-D shape analysis of high-aspect-ratio features using through-focus scanning optical microscopy," Optics Express, 24(15), 16574-16585 (2016).

[19] R. K. Attota, and H. Park, "Optical microscope illumination analysis using through-focus scanning optical microscopy," Opt Lett, 42(12), 2306-2309 (2017).

[20] S.-W. Park, J. H. Lee, and H. Kim, [3D digital holographic semiconductor metrology using Fourier Modal Method] OSA, Fukuoka, Japan(2017).

[21] S. Han, T. Yoshizawa, S. Zhang *et al.*, "Tip/tilt-compensated through-focus scanning optical microscopy," Proc. of SPIE, 10023, 100230P (2016).

[22] M. Ryabko, A. Shchekin, S. Koptyaev *et al.*, "Through-focus scanning optical microscopy (TSOM) considering optical aberrations: practical implementation," Optics Express, 23(25), 32215-32221 (2015).

[23] M. Ryabko, S. Koptyaev, A. Shcherbakov *et al.*, "Motion-free all optical inspection system for nanoscale topology control," Optics Express, 22(12), 14958-14963 (2014).

[24] M. Ryabko, Koptyev, S., Shchekin, A., Medvedev, A., "Improved critical dimension inspection for the

semiconductor industry," SPIE Newsroom, (2014).
[25] V. Vartanian, R. Attota, H. Park *et al.*, "TSV reveal height and dimension metrology by the TSOM method," Proc. of SPIE, 8681, 86812F (2013).
[26] S. Usha, Shashikumar, P.V., Mohankumar, G.C., Rao, S.S., "Through Focus Optical Imaging Technique To Analyze Variations In Nano-Scale Indents," International Journal of Engineering Research & Technology, 2(5), 18 (2013).
[27] M. V. Ryabko, S. N. Koptyaev, A. V. Shcherbakov *et al.*, "Method for optical inspection of nanoscale objects based upon analysis of their defocused images and features of its practical implementation," Optics Express, 21(21), 24483-24489 (2013).
[28] M. Esser, [NIST Nanoscale Dimensioning Technique Wins R&D 100 Award], (2010).
[29] S. I. Association, [The International Technology Roadmap for Semiconductors (ITRS)] Semiconductor Industry Association, San Jose(2016).
[30] S. I. Association, [The International Technology Roadmap for Semiconductors (ITRS)], San Jose(2012).
[31] IEEE, [International Roadmap for Semiconductor Devices and Systems 2017 Edition: Metrology] IEEE, (2018).
[32] [SEMI 3D5-0613 Guide For Metrology Techniques To Be Used In Measurement Of Geometrical Parameters Of Through- Silicon Vias (TSVs) In 3DS-IC Structures] SEMI, (2013).
[33] B. Bunday, T. A. Germer, V. Vartanian *et al.*, "Gaps Analysis for CD Metrology Beyond the 22 nm Node," Metrology, Inspection, and Process Control for Microlithography Xxvii, 8681, (2013).
[34] J. Panyam, and V. Labhasetwar, "Biodegradable nanoparticles for drug and gene delivery to cells and tissue," Advanced drug delivery reviews, (2012).
[35] R. M. Crist, J. H. Grossman, A. K. Patri *et al.*, "Common pitfalls in nanotechnology: lessons learned from NCI's Nanotechnology Characterization Laboratory," Integrative Biology, 5(1), 66-73 (2013).
[36] S. Wang, C. Querner, T. Dadosh *et al.*, "Collective fluorescence enhancement in nanoparticle clusters," Nature Communications, 2, 364 (2011).
[37] H. Kang, M. L. Clarke, S. H. D. P. Lacerda *et al.*, "Multimodal optical studies of single and clustered colloidal quantum dots for the long-term optical property evaluation of quantum dot-based molecular imaging phantoms," Biomedical Optics Express, 3(6), 1312-1325 (2012).
[38] M. Minsky, "Memoir on Inventing the Confocal Scanning Microscope," Scanning, 10(4), 128-138 (1988).
[39] P. Bon, N. Bourg, S. Lecart *et al.*, "Three-dimensional nanometre localization of nanoparticles to enhance super-resolution microscopy," Nature Communications, 6, (2015).
[40] S. Abrahamsson, J. J. Chen, B. Hajj *et al.*, "Fast multicolor 3D imaging using aberration-corrected multifocus microscopy," Nature Methods, 10(1), 60-U80 (2013).
[41] N. Chenouard, I. Smal, F. de Chaumont *et al.*, "Objective comparison of particle tracking methods," Nature Methods, 11(3), 281-U247 (2014).
[42] K. M. Taute, S. Gude, S. J. Tans *et al.*, "High-throughput 3D tracking of bacteria on a standard phase contrast microscope," Nature Communications, 6, (2015).
[43] L. Gardini, M. Capitanio, and F. S. Pavone, "3D tracking of single nanoparticles and quantum dots in living cells by out-of-focus imaging with diffraction pattern recognition," Scientific Reports, 5, (2015).
[44] J. M. Gineste, P. Macko, E. A. Patterson *et al.*, "Three-dimensional automated nanoparticle tracking using Mie scattering in an optical microscope," J Microsc, 243(2), 172-8 (2011).
[45] R. K. Attota, "Step beyond Kohler illumination analysis for far-field quantitative imaging: angular illumination asymmetry (ANILAS) maps," Optics Express, 24(20), 22616-22627 (2016).
[46] R. Attota, and R. Silver, "Optical microscope angular illumination analysis," Optics Express, 20(6), 6693-6702 (2012).
[47] R. Attota, M. Stocker, R. Silver *et al.*, "Through-focus scanning and scatterfield optical methods for advanced overlay target analysis." 7272, 13.

Roles and Responsibilities of NIST in the Development of Documentary Standards

Prem Rachakonda, Bala Muralikrishnan, Daniel Sawyer

Dimensional Metrology Group,
Engineering Physics Division
National Institute of Standards and Technology, Gaithersburg, MD

1 INTRODUCTION

The National Institute of Standards and Technology (NIST[*]) is a non-regulatory federal agency of the Department of Commerce and is the national metrology institute of the US. NIST's role in the development of voluntary consensus standards (VCS) is rooted in many policy decisions and government directives that happened in the 1980s and 1990s. NIST has been a leader in the development of many standards, including documentary[†] standards, ever since its founding in 1901.

The Dimensional Metrology Group (DMG) at NIST develops, performs, and delivers measurements, physical and documentary standards that promote US innovation and industrial competitiveness. NIST DMG has been involved in many standards committees that are a part of International Organization for Standardization (ISO), American Society for Mechanical Engineering (ASME) and more lately ASTM. Recently, DMG, along with various other organizations was involved in the development of an ASTM documentary standard for 3D imaging systems. NIST led the effort and was a major contributor in developing this standard and this activity led to the publication of the ASTM E3125-17 standard in 2017. This standards development process was systematic per the rules and regulations of ASTM, which in turn enabled a balanced approach that addressed the concerns of the stakeholders while maintaining their involvement and establishing a consensus.

This paper describes the basis for NIST's authority to participate and drive the development of documentary standards. The paper will then describe the role of DMG in the standards development process. This includes some of the technical and procedural challenges faced by the working group that published the ASTM E3125-17 3D imaging standard.

2 STANDARDS DEVELOPMENT & ADOPTION BY GOVERNMENT AGENCIES

Many government agencies have been involved in the development of standards to achieve their organizational missions. The Department of Defense (DoD) was known to have over 21,000 government unique standards (GUS) in use prior to the year 1997 [1]. In 1994, the then Secretary of Defense, William J. Perry wrote a policy memorandum, now known as the "Perry Memo"[2][3]. This memorandum highlighted the increasing costs of relying on GUS and directs DoD to use non-government performance standards for its procurement process.

The non-government performance standards mentioned in the "Perry Memo" are voluntary consensus standards (VCS) that are developed by standards development organizations (SDOs)

[*] The names of the organizations mentioned in this paper varied throughout their history, but will be referred in this paper with their current name unless otherwise specified. ASTM International will referred to as ASTM in this paper.

[†] According to ISO, a documentary standard is defined as a document, established by consensus, and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of optimum degree of order in each context.

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

using agreed-upon procedures. The next few sub-sections will detail the history of SDOs, and standards development in the US and how they shaped the events that transpired before and after the "Perry Memo".

2.1 Consolidation of the standards development activity in the US

Prior to 1900s there were many organizations like the American Society of Mechanical Engineers (ASME), ASTM etc. that led the development of standards for the greater good of the society. These professional organizations took up the task of addressing various safety and commercial interests of that era. Soon, it became evident that there was a need to coordinate the standards development activities in the US. In 1918 five engineering societies and three government agencies came together to establish an organization that is known today as the American National Standards Institute (ANSI). The purpose of ANSI's founding was to coordinate standards development, approve national consensus standards and halt user confusion on acceptability. The five engineering societies that founded ANSI were the Institute of Electrical and Electronics Engineers (IEEE), the American Institute of Mining and Metallurgical Engineers (AIME), the American Society of Civil Engineers (ASCE), the ASME, and the ASTM. These five societies, who were themselves core members of the United Engineering Society (UES), subsequently invited the US Departments of War, Navy, and Commerce to join them as founders [4]. The US Departments of War and Navy eventually became the US DoD and US Department of Commerce is the parent organization of NIST.

ANSI was founded as a private non-profit organization and does not develop any standards, but accredits and assesses the competence of the numerous SDOs [5]. ANSI also approves individual documentary standards developed by SDOs and designates them as American National Standards (ANS) [6]. ANS designated standards undergo additional scrutiny that involves accreditation of consensus procedures, neutral oversight, approval process, appeals process, and procedural audit. These additional requirements generally align with the policies set forth by OMB A-119. However, it is up to the SDO to consider designating an approved standard as an ANS by following ANSI's procedures. As of 2015, there were more than 240 ANSI accredited SDOs, over 100,000 VCS published in the US and over 11,000 ANS. For example, the ASME Y14.5M-1994 for "Dimensioning and Tolerancing" is designated as an ANS. ANSI also serves as the US national standards body representative to ISO and many other regional and international standards activities [7,8,9].

There are hundreds of national and international SDOs around the world. Some examples of SDOs accredited by ANSI are ASTM, ASME, IEEE and Underwriter Laboratories (UL). In many countries a centralized body drives the standards activities, which typically is that country's government and is considered a top-down approach. In the US, the private sector drives the standards development activities, which is considered a bottom-up approach. Many SDOs use VCS development processes that ensure due process such as requiring openness, balance, appeals, and consensus.

2.2 US government directives

The US Office of Management and Budget (OMB) issued a circular OMB A-119, in 1980, titled "Federal participation in the development and use of voluntary standards", which was then revised in October 1982. This was considered a groundbreaking policy. The revised circular in 1982 directed federal agencies to participate in SDOs and rely on VCS not just in procurement activities, but also in its regulations. However, the usage of GUS continued and as of 1996 there were still about 34000 GUS used by the federal government [10].

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

Subsequent to the "Perry Memo" in 1994, the policy on usage of VCS for federal agencies was reviewed. This led to the passing of the National Technology Transfer and Advancement Act of 1995 (NTTAA, public law 104-113) [11]. The NTTAA incorporated many of the policy directives of OMB A-119 into a law. OMB A-119 underwent three more revisions since the first revision in 1982 [12]. The 1993 revision [13] reflected the US trade obligations, the 1998 revision [14] was updated to reflect the requirements of the NTTAA and the 2016 revision [15] reflected the changes in numerous federal policies, trade obligations and practices.

The directive for all the federal agencies, including NIST, to participate in the development and adoption of VCS comes from the NTTAA and OMB A-119. The NTTAA mandates that all federal agencies adopt VCS wherever possible, except where inconsistent with law or impractical and avoid the development of GUS. An example of these directives is DoD replacing a government standard, MIL-L-13762 for lead alloy coatings, developed in 1983, with a consensus standard, ASTM-A308 in 1997[16]. GUS are still used by the US federal government for safety, security, and specialized military equipment where appropriate. The primary purpose of NTTAA and OMB A-119 was to eliminate the cost of developing GUS and rely on VCS for such needs wherever possible. As of fiscal year, 2016, there are only about 70 GUS reported in lieu of existing VCS used by federal government agencies (excluding DoD and the National Aeronautics and Space Administration) since reporting began in 1997 [17].

There are many more GUS that are presently in use mostly due to the lack of an equivalent VCS, older GUS that have not been replaced by VCS, or GUS developed due to statutory requirements. For example, the Federal Information Processing Standards (FIPS) published by NIST are government standards that are mandated by law [18], which in its present form address information security related issues in the federal government. Over the years, many FIPS standards were withdrawn [19], and presently only nine FIPS standards are still in use [20]. Another example of usage of GUS by a federal agency is the "Fire protection for shipyards" standard used by the Occupational Safety and Health Administration (OSHA) instead of available VCS. According to OSHA, it was determined that there was no single VCS that covered all the topics for its final ruling [17] and warrants the usage of a GUS.

Though the adoption of VCS is voluntary in nature, VCS can become a law if government agencies adopt them into regulations by a process known as incorporation by reference (IBR). For example, many standards developed by ASTM and UL have been adopted by federal, state, and local governments or codified into law [21]. Many VCS documents are not free of charge and the SDOs may hold the copyrights. However, adoption of VCS into regulations through IBR by federal agencies is permitted as long as the standard is “reasonably available” to the interested parties. The criteria for a VCS being "reasonably available" is detailed in the Office of the Federal Register IBR handbook [22].

2.3 Role of the Department of Commerce and NIST

The US Department of Commerce [23,24] promotes job creation, economic growth, sustainable development, and improved living standards of all Americans by working in partnership with business, universities, and workers. It accomplishes its mission through direct assistance to businesses and communities, targeted investments in world-class research, science, and technology, and through various programs that foster innovation, entrepreneurship, and competitiveness. NIST is one of the 12 bureaus of DOC and its mission is to promote US innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life [25]. NIST has

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

always been at the forefront of development of documentary standards, but it is not an SDO[‡][26]. With the emergence of NTTAA and OMB A-119 directives, NIST was also assigned the role of the coordinating standards and conformity assessment activities and reporting government-wide progress annually to OMB.

2.4 NIST Dimensional Metrology Group activities in documentary standards

The Dimensional Metrology Group at NIST has been leading and supporting the development many documentary standards for over 60 years. DMG works with manufacturers, users, and other stakeholders on committees of various SDOs in this process. Considerable amount of DMG's research and development efforts supports the development of these documentary standards, primarily as part of ISO, ASME and ASTM. NIST DMG staff also hold leadership positions in about 20 standards committees that are a part of various SDOs like ASME, ISO and ASTM. Some of the recent examples of documentary standards led by DMG were the ASME B89 series, ASTM E57 and ISO 10360 series of standards. Examples of documentary standards from these activities include ASME B89.4.19-2006[27], and ISO 10360-10:2016[28] standard for laser trackers and ASTM E3125-17[29] standard for laser scanners.

3 BACKGROUND OF THE ASTM E3125-17 3D IMAGING STANDARD

The ASTM E3125-17 standard describes the performance evaluation methods for a class of 3D imaging systems called Terrestrial Laser Scanners (TLSs). An overview of the ASTM E3125-17 standard was reported by Rachakonda et. al [30]. TLSs are used in a variety of applications including reverse engineering, surveying, large scale assembly and historical artifact preservation. TLSs were investigated at NIST for use in a variety of applications starting in the late 1990s. NIST recognized the potential impact of these instruments in a range of industrial applications and that standards were needed for these class of instruments. Therefore, NIST organized three workshops between 2003 and 2006 which brought together instrument manufacturers, end users and other organizations to determine the needs and interests of all the stake holders [31,32]. The participants of these workshops agreed that the development of documentary standards would benefit the TLS users and would help promote widespread use of the technology.

The process of selection of ASTM as an SDO for 3D imaging standards was also achieved through a consensus process by participants of the NIST workshop in 2006 [32]. Several SDOs were initially considered and the choices were narrowed down to two SDOs. These SDOs were sent questionnaires about their standards development process. The responses were then sent to the participants of the workshop and were asked to vote for an SDO. A clear majority of the participants selected ASTM as the SDO. After this workshop, the ASTM E57 committee was formally established in 2006 to address issues related to 3D imaging systems such as TLSs, optical range cameras etc. [33]. This committee presently has multiple sub-committees as illustrated in *Figure 1*. Each sub-committee may have multiple work items, which when successful are published as documentary standards. As of the writing of this paper, there are eight published standards under E57 including five standards under E57.02 sub-committee on Test Methods [34].

[‡] Only one laboratory of NIST, the Information Technology Laboratory(NIST/ITL), is accredited by ANSI as an SDO and primarily retains this status due to various historical reasons that precede NTTAA.

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

After establishing the ASTM E57 committee, a working group started work on evaluating the ranging capability of 3D imaging systems, as this was the fundamental measurement of these systems. In 2015, this working group published the ASTM E2938-15 standard for 3D imaging systems to evaluate relative-range [35]. The scope of this standard was limited to the relative-range as the main source of error was from the range measurement of the instrument.

In 2013, another working group was convened under ASTM E57 (WK43218), to address the volumetric performance of the TLS. This was since, both range and angular errors contribute to the instrument errors when any point-to-point distance is measured. This effort was led by DMG staff as they had considerable experience on laser trackers due to their efforts towards ASME B89.4.19-2006[27] standard. Laser trackers and TLSs are very similar in construction and both these instruments were studied extensively by Muralikrishnan et.al. [36,37].

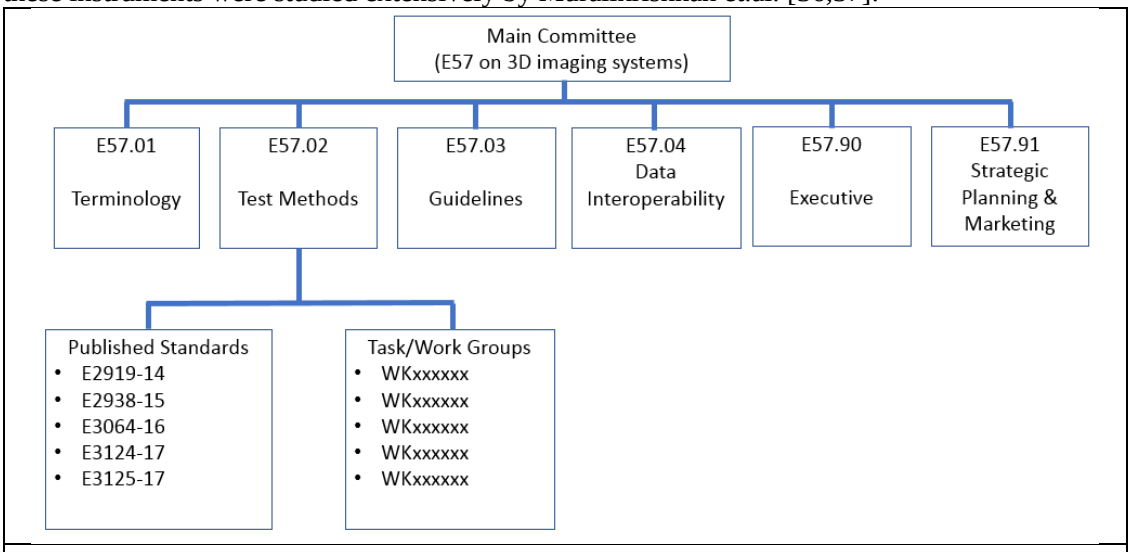


Figure 1: ASTM organization structure showing published standards under E57.02

NIST role in this standards development process was both as a facilitator as well as to provide subject matter expertise on TLSs and dimensional metrology instrumentation. In these roles, NIST could not exert any more authority than other stakeholders in influencing the direction of the standard. All the decisions that were taken during this process had to be agreeable to all the stakeholders and it was NIST's responsibility to facilitate consensus. NIST DMG staff worked through both technical and procedural hurdles to achieve this task. As part of the ASTM sub-committee, many of the guidelines and procedures set forth by ASTM were consulted to handle many of the issues that cropped up during this process. Some of the issues encountered by the WK43218 working group are described next.

3.1 Stakeholders

Many of the stakeholders that constituted WK43218 were a part the working group that published ASTM E2938-15. Their objectives remained the same – to standardize the performance evaluation of TLSs, but in the entire work volume of the instrument. NIST publicized the activities of this working group at various venues, websites, conferences, symposia, and invited participants to contribute to this effort. The participants of this working group consisted of national metrology institutes, instrument manufacturers, distributors, end users and other subject matter experts.

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

3.2 Technical challenges

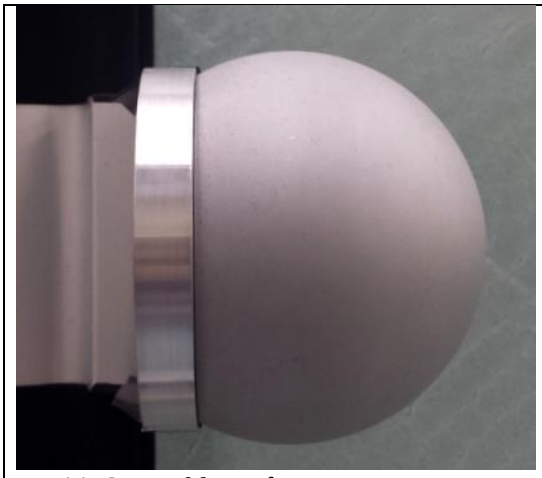
There were many technical challenges during the development of the ASTM E3125-17 standard. NIST DMG was uniquely positioned to address these challenges due to its technical expertise, facilities, and resources. Two of the notable hurdles were the availability of artifacts and algorithms to process TLS data. If not chosen appropriately, both the artifact and the algorithms could introduce an error in the performance metrics that are not representative of the instrument.

Spheres were chosen as ideal targets for this activity for reasons that are detailed in [38,39]. The challenge was the availability of commercial spheres that were suitable to be measured both by the instrument under test and the reference instrument. Early on, it was decided that the test procedures at NIST would be using a laser tracker as the reference instrument. Measuring large reference lengths was challenging on stationary CMMs and laser trackers provided a way to perform in-situ measurements. A variety of artifacts were considered by the ASTM working group and the breakthrough came with the availability of a commercial artifact known as "Integration sphere" depicted in Figure 2. This sphere had a circularity of < 10 µm and it was designed to be measured by both the TLS and the laser tracker.

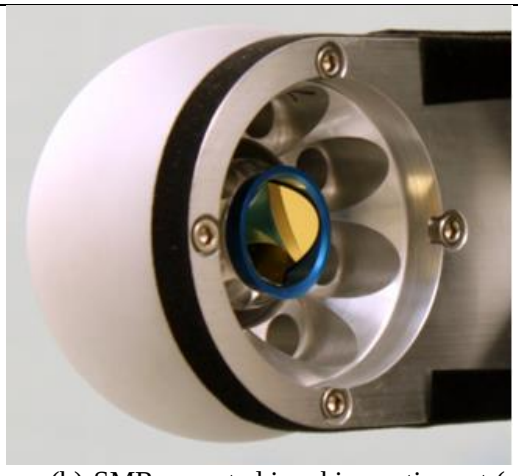
Algorithms to process the TLS data were the second major challenge. As increasing amount of data was being collected, it was observed sphere datasets were littered with a lot of phantom or spurious points that did not belong to the sphere. The location of these spurious points varied based on the instrument and its ranging technology and had to be excluded. A variety of algorithms were considered for this activity and an algorithm known as the "Cone-Cylinder algorithm" was developed for this activity and this algorithm resulted in minimal number of spurious points. The details of these algorithms were reported by Rachakonda et. al [40].

Following are some of the other technical challenges that were encountered by this working group:

1. Instrument downtime, due to repair, servicing, or calibration
2. Proprietary nature of instrument's construction, data format etc.
3. Lack of resources, facilities, equipment, software, and artifacts to independently examine the technical issues by all the stakeholders.



(a) Scannable surface



(b) SMR mounted in a kinematic nest (not visible) at the sphere center

Figure 2: Commercial sphere used on the grid

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

3.3 Procedural challenges

One of the major procedural challenge happened at the start of this activity and it was to define the scope of the standard. This is one of the most common challenges encountered by many working groups that develop documentary standards. The ASTM E2938-15 standard can be used to evaluate any 3D imaging system, however, ASTM E3125-17 limits the standard only for 3D imaging systems that have a spherical coordinate system such as TLSs. This was because, the TLS error sources were well understood prior to the start of the standards activity [36,37]. Based on this work, the scope was limited to using test positions that were sensitive to the instrument error sources. The scope was also limited to the usage of reference lengths instead of consistency checks and the usage of targets that have optical and mechanical properties suitable for a TLS to scan the target in its test volume.

A significant procedural challenge that was encountered during this activity was new stakeholders participating in the midst of development of the standard and stakeholders participating on an intermittent basis. Concerns about the direction of the standard and the requisite time commitment are a couple of factors that affect stakeholder participation. A considerable amount of effort was expended to satisfy stakeholders about the direction of the standard.

One other challenge was negative votes from non-participating members of the ASTM E57 subcommittee. For a standard to be approved, it must be voted on by the subcommittee members. ASTM requires responses from about 60 % of the subcommittee members and 90 % of those responses must be in favor for the standard to be approved. All the negative votes had to be addressed per ASTM procedures. The negative votes were discussed at length to determine if they were persuasive or not and technical responses were drafted. The recommended changes to proposed standard were then submitted to the subcommittee for approval. Once approved, the standard had to be re-balloted as some of the changes were substantial and not editorial.

Following are some of the other procedural challenges that were encountered by this working group:

1. Lack of sustained interest from subject matter experts.
2. Long lead times for procurement, instrument repairs and manuscript publication.
3. Organizational budgetary cycles not matching the standards development life cycle.

4 SUMMARY

NIST has historically been a leader in supporting the development of many documentary standards. As new market driven requirements emerge NIST provides its expertise to both the private industry and the US government in the development of VCS. In this context, this paper describes NIST's role which is defined by its statutory authority and further set out in the NTTAA and OMB A-119. The paper then describes some of the challenges faced during the development of one VCS, ASTM E3125-17, that was developed according ASTM rules and regulations.

5 DISCLAIMER

Commercial equipment and materials may be identified to specify certain procedures. In no case does such identification imply recommendation or endorsement by the NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

6 ACKNOWLEDGEMENTS

The authors would like to thank all the participants of the ASTM WK43218 whose contributions made the ASTM E3125-17 a robust documentary standard for TLSs. The authors

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

would also like to thank Geraldine Cheok of the NIST Engineering Laboratory, David Alderman, Mary Donaldson & Warren Merkel of the NIST Standards Coordination Office and Michael Hogan of NIST ITL for their helpful discussions, document review and feedback regarding voluntary consensus standards.

7 REFERENCES

[1] M. Donaldson, N. Rioux, "Fifteenth Annual Report on Federal Agency Use of Voluntary Consensus Standards and Conformity Assessment", NISTIR 7857 https://standards.gov/nttaa/resources/nttaa_ar_2011.pdf (Accessed on 6/6/2018)

[2] W. Perry, "A New Way of Doing Business", June 1994 https://www.sae.org/standardsdev/military/milperry.htm (Accessed on 6/6/2018)

[3] J. Gansler, W. Lucyshyn, "Commercial-off-the-shelf (COTS): Doing it Right", http://www.dtic.mil/dtic/tr/fulltext/u2/a494143.pdf (Accessed 6/6/2018)

[4] History of the American National Standards Institute: https://www.ansi.org/about_ansi/introduction/history?menuid=1 (Accessed on 6/6/2018)

[5] Overview of the U.S. Standardization System, Second Edition, 2007, American National Standards Institute https://share.ansi.org/Shared%20Documents/News%20and%20Publications/Brochures/U.S.%20Standardization%20System-07.pdf (Accessed 6/6/2018)

[6] ANSI Essential Requirements: Due process requirements for American National Standards https://www.ansi.org/essentialrequirements/ (Accessed on 6/6/2018)

[7] ISO Programs – Overview: https://www.ansi.org/standards_activities/iso_programs/overview (Accessed on 6/6/2018)

[8] NISTIR 8007: A Review of USA Participation in ISO and IEC http://dx.doi.org/10.6028/NIST.IR.8007 (Accessed on 6/6/2018)

[9] MOU between ANSI and NIST https://share.ansi.org/Shared%20Documents/About%20ANSI/Memoranda%20of%20Understanding/ansinist_mou.pdf (Accessed 6/6/2018)

[10] NIST SP 806, 1996 edition: Standards Activities of Organizations in the United States https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication806e1996.pdf (Accessed 6/6/2018)

[11] National Technology Transfer and Advancement Act of 1995 https://www.nist.gov/standardsgov/national-technology-transfer-and-advancement-act-1995

[12] OMB Circular A-119: Federal Participation in the Development and Use of Voluntary Standards (1982 revision):

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

[13] OMB Circular A-119: Federal Participation in the Development and Use of Voluntary Standards (1993 revision) https://clintonwhitehouse1.archives.gov/White_House/EOP/OMB/html/circulars/a119/a119.html (Accessed 6/6/2018)

[14] OMB Circular A-119: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities (1998 revision) https://www.whitehouse.gov/wp-content/uploads/2017/11/Circular-119-1.pdf (Accessed 6/6/2018)

[15] OMB Circular A-119: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities (2016 revision): https://www.nist.gov/sites/default/files/revised_circular_a-119_as_of_01-22-2016.pdf (Accessed on 6/6/2018)

[16] Department of Defense Index of Specifications and Standards. Part 2. Numerical Listing http://www.dtic.mil/docs/citations/ADA273295 (Accessed on 6/6/2018)

[17] FY 2016 Government Unique Standards used in lieu of Voluntary Consensus Standards https://standards.gov/NTTAA/resources/FY2016_GUSs_used_in_lieu_of_VCSs.pdf (Accessed on 6/6/2018)

[18] FIPS Pub 199: Standards for Security Categorization of Federal Information and Information Systems. https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.199.pdf (Accessed on 6/6/2018)

[19] Announcing Approval of the Withdrawal of Ten Federal Information Processing Standards https://csrc.nist.gov/News/2008/Announcing-Approval-of-the-Withdrawal-of-Ten-FIP-S (Accessed on 6/6/2018)

[20] Current FIPS https://www.nist.gov/itl/current-fips (Accessed on 6/6/2018)

[21] Safety and Health Regulations for Longshoring – Personal Protective Equipment https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10493 (Accessed on 6/6/2018)

[22] IBR Handbook, 2017 https://www.archives.gov/files/federal-register/write/handbook/ibr.pdf (Accessed on 6/6/2018)

[23] The Commerce Mission Statement: http://www.osec.doc.gov/bmi/budget/strtgc/Aintro.pdf (Accessed on 6/6/2018)

[24] The Department of Commerce Budget in Brief: http://osec.doc.gov/bmi/Budget/FY14BIB/ENTIREBIB.pdf (Accessed on 6/6/2018)

[25] NIST Mission, Vision, Core Competencies, and Core Values https://www.nist.gov/about-nist/our-organization/mission-vision-values (Accessed on 6/6/2018)

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

[26] ANSI Accreditation of ITL: https://www.nist.gov/itl/ansi-accreditation-itl (Accessed on 6/6/2018)

[27]Performance Evaluation of Laser-Based Spherical Coordinate Measurement Systems https://www.asme.org/products/codes-standards/b89419-2006-performance-evaluation-laserbased (Accessed on 6/6/2018)

[28] ISO 10360-10:2016: Geometrical product specifications (GPS) -- Acceptance and reverification tests for coordinate measuring systems (CMS) -- Part 10: Laser trackers for measuring point-to-point distances https://www.iso.org/standard/56662.html (Accessed on 6/6/2018)

[29] ASTM E3125-17 Standard Test Method for Evaluating the Point-to-Point Distance Measurement Performance of Spherical Coordinate 3D Imaging Systems in the Medium Range, ASTM International, West Conshohocken, PA, 2017, https://doi.org/10.1520/E3125-17 (Accessed on 6/6/2018)

[30] P. Rachakonda, B. Muralikrishnan, K. Shilling, D. Sawyer, G. Cheok, "An Overview of Activities at NIST Towards the Proposed ASTM E57 3D Imaging System Point-to-point Distance Standard", Journal of the CMSC, October 2017.

[31] Cheok, G., "NISTIR 7266: Proceedings of the 2nd NIST LADAR performance evaluation workshop – March 15 - 16, 2005" https://dx.doi.org/10.6028/NIST.IR.7266 (Accessed on 6/6/2018)

[32] Cheok, G., "NISTIR 7357: Proceedings of the 3rd NIST workshop on the performance evaluation of 3D imaging systems – March 2 - 3, 2006" https://dx.doi.org/10.6028/NIST.IR.7357 (Accessed on 6/6/2018)

[33] ASTM E57 Committee on 3D Imaging Systems https://www.astm.org/COMMITTEE/E57.htm (Accessed on 5/18/2018)

[34] ASTM Subcommittee E57.02 on Test Methods https://www.astm.org/COMMIT/SUBCOMMIT/E5702.htm (Accessed on 5/18/2018)

[35] ASTM E2938-15 Standard Test Method for Evaluating the Relative-Range Measurement Performance of 3D Imaging Systems in the Medium Range, ASTM International, West Conshohocken, PA, 2015, https://doi.org/10.1520/E2938-15 (Accessed on 6/6/2018)

[36] B. Muralikrishnan, D. Sawyer, et al., "ASME B89.4.19 Performance Evaluation Tests and Geometric Misalignments in Laser Trackers" Journal of Research (NIST), 114, 21-35 (2009)

[37] B. Muralikrishnan, M. Ferrucci, D. Sawyer, et al., "Volumetric performance evaluation of a laser scanner based on geometric error model", Precision Engineering, 40, p. 139–150, 2015.

[38] Muralikrishnan, B., Shilling, M., Rachakonda, P., Ren, W., Lee, V., Sawyer, D., "Toward the development of a documentary standard for derived-point to derived-point distance performance evaluation of spherical coordinate 3D imaging systems", Journal of Manufacturing Systems, Vol. 37, Part 2, October 2015 pp550-557

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

[39] Rachakonda, P., Muralikrishnan, B., Shilling, M., Cheok, G., Lee, V., Blackburn, C., Everett, D., Sawyer, D., "Targets for relative range error measurement of 3D imaging systems", Journal of the CMSC, Vol. 12, No. 1, Spring 2017

[40] Rachakonda P, Muralikrishnan B, Cournoyer L, Cheok G, Lee V, Shilling K, Sawyer D (2017) Methods and considerations to determine sphere center from terrestrial laser scanner point cloud data. *Measurement Science and Technology* 28(10):105001. https://doi.org/10.1088/1361-6501/aa8011 (Accessed on 6/6/2018)

Muralikrishnan, Balasubramanian; Rachakonda, Prem; Sawyer, Daniel.
"Roles and Responsibilities of NIST in the Development of Documentary Standards."
Paper presented at The Coordinate Metrology Society Conference, Reno, NV, United States. July 23, 2018 - July 27, 2018.

Proceedings of the ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
IDETC2018
August 26-29, 2018, Quebec City, Quebec, Canada

DETC2018-85109

On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards

Craig M. Shakarji
Physical Measurement Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.
craig.shakarji@nist.gov

Vijay Srinivasan
Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.
vijay.srinivasan@nist.gov

Abstract

Constrained least-squares fitting has gained considerable popularity among national and international standards committees as the default method for establishing datums on manufactured parts. This has resulted in the emergence of several interesting and urgent problems in computational coordinate metrology. Among them is the problem of fitting inscribing and circumscribing circles (in two-dimensions) and spheres (in three-dimensions) using constrained least-squares criterion to a set of points that are usually described as a 'point-cloud.' This paper builds on earlier theoretical work, and provides practical algorithms and heuristics to compute such circles and spheres. Representative codes that implement these algorithms and heuristics are also given to encourage industrial use and rapid adoption of the emerging standards.

1. Introduction

Recent years have seen the emergence of constrained least-squares as the preferred criterion for establishing digital datums in ASME and ISO (International Organization for Standardization) standards committees [1]. There is also a nascent interest in applying constrained least-squares fitting for tolerancing purposes beyond datums. Unconstrained least-squares fitting has been used in digital metrology for decades, and algorithms have been published to implement them [2-5]. Now there is industrial demand for constrained least-squares fitting algorithms and guidance for their implementations.

An algorithm must theoretically guarantee the correct solution to a problem, but it may not be ultra-efficient in execution time or memory space. On the other hand, heuristics may find the correct solution to the problem, but they cannot theoretically guarantee the correctness. Heuristics are often simpler to implement and they may compute quickly.

The general problem of constrained least-squares fitting has attracted the attention of several researchers in several domains [6-8]. In the field of computational coordinate metrology, preliminary algorithms for constrained least-squares fitting of planes, parallel planes, and intersecting planes for establishing digital datums have been published only recently [9-11], and these algorithms have been implemented in commercial software. This paper builds on recent theoretical results [12] for constrained least-squares fitting of circles and sphere, and provides algorithms for computing them. It also provides heuristics to compute such circles and spheres. Representative codes are provided to implement these algorithms and heuristics. These algorithms, heuristics, and codes are the major contributions of this paper.

The rest of the paper is organized as follows. Section 2 analyzes the nature of the optimization problem. Algorithms to compute the global optimum solution are provided in Section 3, along with codes to implement one of these algorithms. Section 4 provides some heuristics and codes to compute such circles and spheres. Opportunities for improvements and extensions are discussed in Section 5. Finally, Section 6 summarizes the results of the paper and offers some directions for future research and development.

2. Nature of the Optimization Problem

Consider a set of n points $\{\boldsymbol{p_1}, \ldots, \boldsymbol{p_n}\}$ in a plane (for fitting a circle) or in space (for fitting a sphere). These are the input points for the optimization problems. Let $\boldsymbol{c}$ be the center and $\boldsymbol{r_c}$ be the radius of a circle or sphere of interest. Also let $\boldsymbol{r_i}$ be the distance between $\boldsymbol{c}$ and $\boldsymbol{p_i}$. Figure 1 illustrates these notations in

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

a plane involving a circle. A similar figure can be imagined for points in space involving a sphere.

For the sake of computational coordinate metrology discussed in this paper, each point will be assumed to have Cartesian coordinates as $\boldsymbol{p_i} = (x_i, y_i)$ in a plane, or $\boldsymbol{p_i} = (x_i, y_i, z_i)$ in space. The center will have its coordinates as $\boldsymbol{c} = (x_c, y_c)$ in the plane, or $\boldsymbol{c} = (x_c, y_c, z_c)$ in space.

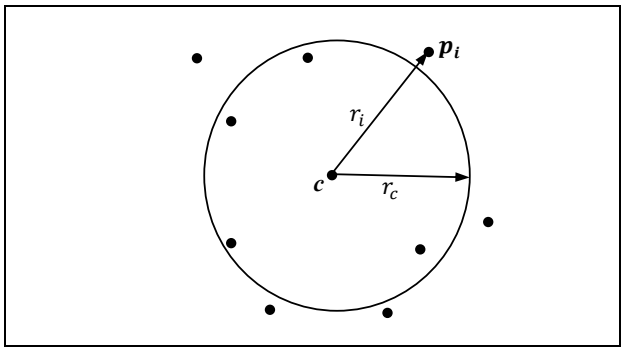


Figure 1. Illustration of notations for the optimization problem.

Following notations established earlier [12], four constrained least-squares fitting problems will be denoted as:

- CL_2IC: Constrained least-squares inscribing circle.
- CL_2CC: Constrained least-squares circumscribing circle.
- CL_2IS_p: Constrained least-squares inscribing sphere.
- CL_2CS_p: Constrained least-squares circumscribing sphere.

With these notations, constrained least-squares fitting of circles and spheres can be posed compactly as the following four constrained optimization problems.

$$\min_{\boldsymbol{c}, r_c} \left[\frac{1}{n}\sum_{i=1}^{n}(r_i - r_c)^2\right]^{1/2}$$

subject to (1)

$$r_c \le r_i, \forall i \text{ for } CL_2IC \text{ and } CL_2IS_p,$$
$$r_c \ge r_i, \forall i \text{ for } CL_2CC \text{ and } CL_2CS_p.$$

The objective function in Eq. (1) is represented in the L_2-norm. It is also commonly referred to as the root-mean-square (RMS) of the deviation. This explains why all the designations of the fitting problems have the CL_2 prefix, denoting that each is a constrained least-squares fit. The same optimum solution can be obtained if the objective function is retained in a simpler form as in $\min_{\boldsymbol{c}, r_c} \sum_{i=1}^{n}(r_i - r_c)^2$ with the same constraints, as done in earlier literature [12].

2.1 Removing the constraints

The objective function in Eq. (1) is a continuous and smooth function (that is, having continuous derivatives) of the coordinates of the center $\boldsymbol{c}$ and the radius r_c. However, the inequality constraints and the non-linear nature of the objective function render the optimization problem more difficult to solve than a corresponding unconstrained optimization problem. The problem can be converted to a more tractable, unconstrained optimization problem by recasting Eq. (1) as

$$\min_{\boldsymbol{c}} \left[\frac{1}{n}\sum_{i=1}^{n}(r_i - r_{min})^2\right]^{1/2} \text{where } r_{\min} = \min_i(r_i)$$
$$\text{for } CL_2IC \text{ and } CL_2IS_p, \text{ and}$$
$$\min_{\boldsymbol{c}} \left[\frac{1}{n}\sum_{i=1}^{n}(r_i - r_{max})^2\right]^{1/2} \text{where } r_{\max} = \max_i(r_i) \quad (2)$$
$$\text{for } CL_2CC \text{ and } CL_2CS_p.$$

The constraints in Eq. (1) have been removed in Eq. (2), and the free variables are reduced to the center coordinates (that is, coordinates of $\boldsymbol{c}$) alone. However, this had been achieved at the expense of some lack of smoothness of the objective functions in Eq. (2), which are now continuous in the center coordinates but with discontinuous gradient vectors.

The non-smooth nature of the objective functions of Eq. (2) can be illustrated with a simple, but generally representative, example. Consider the following example problem with only four points in a plane:

Example E1. $\boldsymbol{p_1}$: $(1.2, 0.9)$, $\boldsymbol{p_2}$: $(-0.95, 1.05)$, $\boldsymbol{p_3}$: $(-1.1, -1.2)$, and $\boldsymbol{p_4}$: $(0.9, -0.8)$.

The objective function of Eq. (2) for the inscribing circles (CL_2IC) for Example E1 is shown in Fig. 2. A clearer illustration of the objective function is shown in Fig.3 as a contour plot, where the input points of Example E1 are marked as '+'.

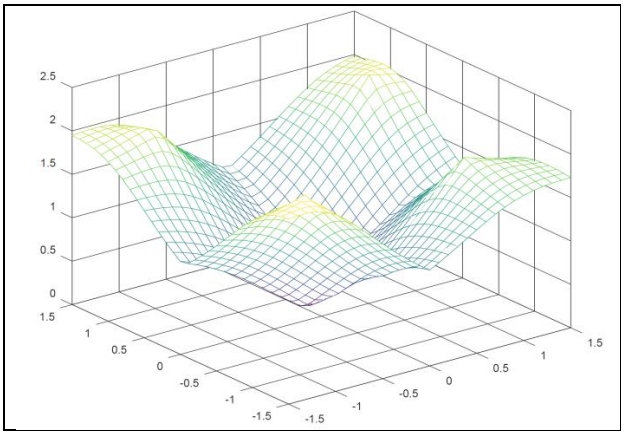


Figure 2. Objective function for Example E1.

It can be clearly seen that most contours in Fig. 3 suffer tangent discontinuities at discrete points. These are the places where the gradient vector of the objective function also suffers discontinuities. The minimum for the objective function in Fig.

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

2 occurs at the bottom of the valley within the inner most contour of Fig. 3 near the origin (0,0).

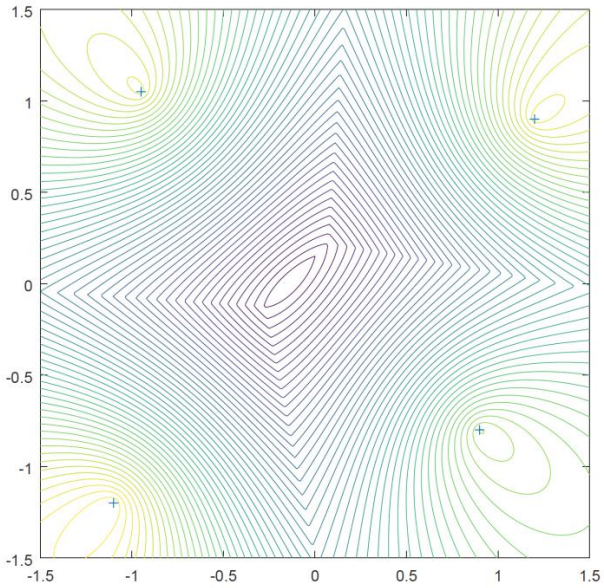


Figure 3. Contour plot of the objective function of Fig. 2.

2.2 Connection to Voronoi diagrams

An interesting and important connection between the behavior of the objective functions of Eq. (2) and the Voronoi diagrams of the input points has been established in a recent theoretical investigation of the optimality conditions for constrained least-squares fitting of circles and spheres [12]. In that investigation, the following theorem was proved.

Theorem 1: The constrained least-squares inscribing and circumscribing circles and spheres must contact at least two input points.

It immediately implies the following theorem.

Theorem 2: The centers of the constrained least-squares circles and spheres must lie on the Voronoi diagram of the input points. More specifically, the center of CL_2IC and CL_2IS_p must lie on the nearest-neighbor Voronoi diagram of the input points; similarly, the center of CL_2CC and CL_2CS_p must lie on the furthest-neighbor Voronoi diagram of the input points.

The connection between the objective functions of Eq. (2) and Voronoi diagrams can be illustrated using Example E1. Figure 4 shows the nearest-neighbor Voronoi diagram for the input points of Example E1. Figure 5 is a superposition of the contour plot (Fig. 3) and the Voronoi diagram (Fig. 4). It can be clearly seen in Fig.5 that the Voronoi edges of Fig. 4 are perfectly aligned with the points of tangent discontinuities of Fig. 3.

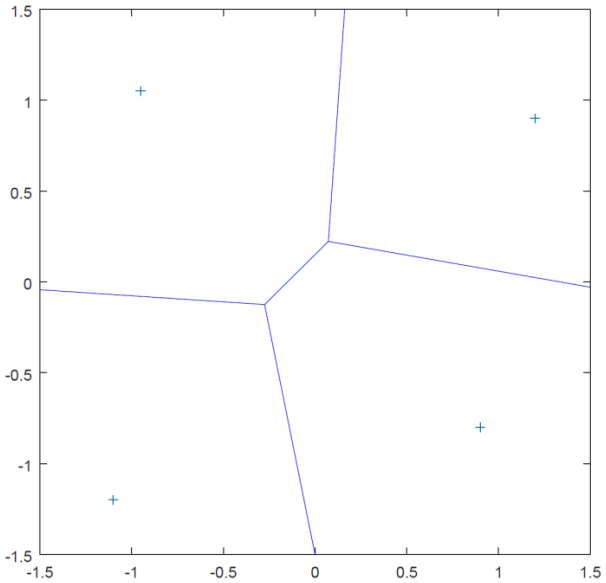


Figure 4. Voronoi diagram of input points in Example E1.

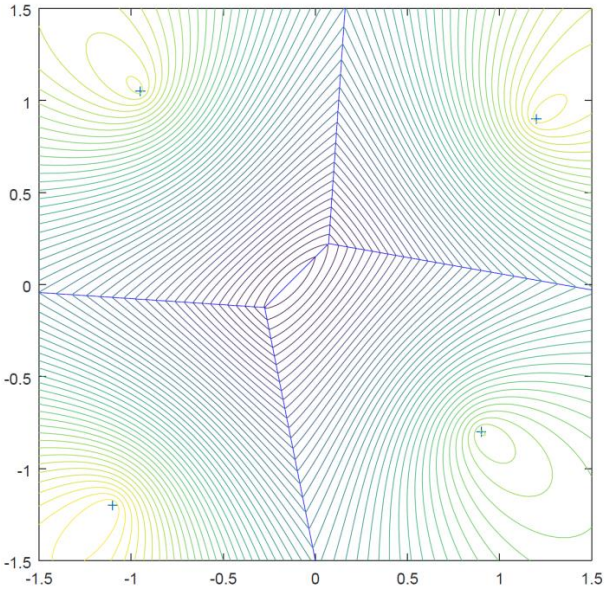


Figure 5. Superposition of Figs. 3 and 4.

The nature of the optimization problem can now be summarized as follows.

- The minimum occurs on the Voronoi diagrams. This fact will be exploited in Section 3 to design and implement algorithms that find the global minimum.
- The minimum occurs where the gradient vector is discontinuous. This fact will be considered in Section 4 in

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

the design and implementation of heuristics that find a minimum.

3. Algorithms and their Implementations

Theorem 2 guarantees that the global minimum to the optimization problems of Eq. (2) can be found by restricting the search to Voronoi diagrams. This cuts the feasible region down by one full dimension. For circles in a plane, the search for the optimal center needs to look no further than a finite number of line segments. Similarly, for spheres in space, the search for the optimal center can be confined to a finite number of convex polygonal regions. Moreover, in both cases, the objective function is smooth (continuous function with continuous derivatives) over each Voronoi edge (in the plane) and over each Voronoi face (in space). These notions will now be described in some detail.

The nearest-neighbor or furthest-neighbor Voronoi diagram for a discrete set of points in a plane is a connected, planar graph $G = (E, V)$ with edge set E and vertex set V [13]. The Voronoi edges are straight line segments. Figure 4 is an example of such a graph. Some of the edges in E will extend to infinity in one direction; these are usually trimmed by a generously large bounding box for computational and display purposes. For a set of n input points in a plane, its Voronoi diagram has $O(n)$ edges and $O(n)$ vertices.

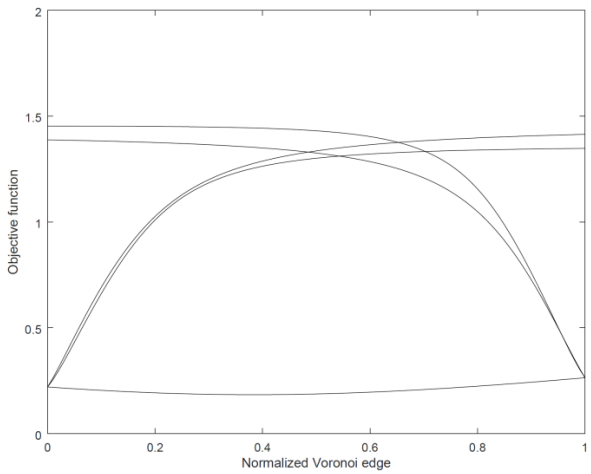


Figure 6. Objective function along the five Voronoi edges shown in Figs. 4 and 5. The bottom-most function is along the only short, finite Voronoi edge. The other four functions are along the half-line Voronoi edges that have been trimmed.

The nearest-neighbor or furthest-neighbor Voronoi diagram for a discrete set of points in (3D) space is also a connected (but not planar) graph $G = (F, E, V)$ with face set F, edge set E, and vertex set V. The faces in F are convex polygons, and the edges in E are straight line segments. Some of the faces and edges will extend to infinity, and they are usually trimmed by a generously large three-dimensional bounding box for computational and display purposes. For a set of n input points in space, the size of its Voronoi diagram can vary quadratically with n.

Even though the gradient vectors of the objective functions of Eq. (2) are not continuous in the region of interest in the plane and in space, the objective functions are locally smooth and continuous in the interior of the Voronoi edges and Voronoi faces. This means that the gradient vector is also continuous along the Voronoi edges and over the Voronoi faces. Figure 6 illustrates the local smoothness of the objective function along each of the five Voronoi edges shown in Figs. 4 and 5. Such local smoothness will be exploited in the design of algorithms in Section 3.1, and in their implementation in Section 3.2.

3.1 Design and analysis of algorithms

A general algorithm to find the global minimum for the problems in Eq. (2) is given in Table 1 in two major steps.

Table 1. Algorithm for circles and spheres.

Algorithm A_CL2x (x stands for IC, CC, ISp, or CSp)
Input: Set of n points, either in a plane or in space.
Output: Center of optimum circle/sphere, its radius, and the minimized objective function.

1. Compute the Voronoi diagram of the input points.
   a. For CL_2IC and CL_2IS_p, it is the nearest-neighbor Voronoi diagram. For CL2CC and CL2CSp, it is the furthest-neighbor Voronoi diagram.
   b. For CL_2IC and CL_2CC, the Voronoi diagram is a connected, planar graph $G = (E, V)$. For CL_2IS_p and CL_2CS_p, it a connected graph $G = (F, E, V)$.
2. Search through the Voronoi graph to find the global minimum.
   a. For CL_2IC and CL_2CC, search along each closed edge in the edge set E to find the minimum, and then find the minimum of all such minima over E. If more than one global minimum is found, return all the global minimum results.
   b. For CL_2IS_p and CL_2CS_p, search over each closed face in the face set F to find the minimum, and then find the minimum of all such minima over F. If more than one global minimum is found, return all the global minimum results.

The correctness of the A_CL2x algorithm is guaranteed by Theorem 2. An analysis of the A_CL2x algorithm requires the consideration of both the steps.

1. The time and space complexity of Step 1 is completely determined by the complexity of computing the Voronoi diagram. One of the most popular algorithm to compute Voronoi diagrams is called `qhull` [14], which can be computed, on average, in $O(n \log n)$ time.
2. The complexity of Step 2 is dominated by the number of Voronoi edges or Voronoi faces; the time required for finding the minimum along each Voronoi edge or Voronoi

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

face is not a function of the input size n, and so it can be considered to be a constant for the purpose of analysis. There are only $O(n)$ Voronoi edges in two-dimensional problems. But there could be $O(n^2)$ Voronoi faces in three-dimensional problems.

Table 2. Code for A_CL2IC

filename: A_CL2IC.m

```
function [c,rc,f]=A_CL2IC(M)
%
% Implementation of an algorithm
%   to fit a constrained (inscribed) least-squares circle in a plane
% Input:
%   M: an nx2 matrix of x and y coordinates of n points
% Output:
%   c:  center of the circle; a 1x2 array of x and y coordinates
%   rc: radius of the circle; a scalar
%   f:  minimized objective function; a scalar
%
% Needs:
%   fminbnd: a built-in function to search for minimum over a Voronoi edge
%   CL2IC_ObjFun: a local function needed by fminbnd
%
  global P % Make the coordinate matrix a global variable in this file
  global psedge % Make starting endpoint of an edge global
  global pfedge % Make finishing endpoint of an edge global
  P = M; % Copy M to P
  [vx, vy] = voronoi(P(:,1), P(:,2)); % Compute (nearest-neighbor) Voronoi diagram
  nedge = size(vx)(2); % Number of edges in the Voronoi diagram
  ps = [vx(1,:); vy(1,:)]; % Vector of starting endpoints of Voronoi edges
  pf = [vx(2,:); vy(2,:)]; % Vector of finishin endpoints of Voronoi edges
  for j=1:nedge % Search over all Voronoi edges
    psedge = ps(:,j); % Starting endpoint of Voronoi edge
    pfedge = pf(:,j); % Finishing endpoint of Voronoi edge
    [tval, fval] = fminbnd(@CL2IC_ObjFun, 0, 1); % Call the built-in minimizer
    cen(:,j) = (1-tval)*psedge + tval*pfedge; % Center coordinates of local minima
    fminval(j) = fval; % Vector of local minima
    end
  [f, ednum] = min(fminval); % Global minimum
  c = cen(:,ednum)'; % Center coordinates of global minimum
  rc = min(sqrt((P(:,1)-c(1)).^2 + (P(:,2)-c(2)).^2)); % Radius at global minimum
end
%
  function f=CL2IC_ObjFun(t)
%   Objctive function needed by fminbnd
%   Input:
%     t: argument (parameter) for the objective function; a scalar
%   Output:
%     f: Objective function = sqaure root of the mean of the squares (RMS)
%        of the deviations of the points from the circle; a scalar
%
    global P % a gloabl variable to get the coordinate matrix
    global psedge % Make starting endpoint of an edge global
    global pfedge % Make finishing endpoint of an edge global
    x = (1-t)*psedge + t*pfedge; % From parameter to coordinates
    r = sqrt((P(:,1)-x(1)).^2 + (P(:,2)-x(2)).^2); % radius vector
    f = (norm(r-min(r)))/sqrt(size(r)(1)); % RMS of deviations from the circle
  end
```


Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

3.2 Implementation of algorithms

Any implementation of an algorithm requires a language to be chosen and some compromises to be made. Table 2 exhibits an implementation of the A_CL2IC algorithm in GNU Octave language [15]. GNU Octave is a free software distributed under the GNU General Public License. It is a 'MATLAB-clone,' which means in practice that GNU Octave code should run under MATLAB, but not necessarily the other way around. GNU Octave has the advantage of free availability and openness, but it is only an interpretive language with some syntactic restrictions (e.g., no support for nested functions).

The A_CL2IC code in Table 2 is generously commented to be almost self-documented. Step 1 of Algorithm A_CL2IC is executed in line 20 using a built-in `voronoi` function of GNU Octave. The rest of the code implements Step 2 of the algorithm. As each Voronoi edge is examined in the `for` loop in lines 24 through 30, a built-in minimizer `fminbnd` is called in line 27 to find the minimum of the objective function along that edge. Here it is assumed that the univariate objective function is smooth with continuous directional derivative along each Voronoi edge, as illustrated in Fig. 6.

The smoothness of the objective function with continuous directional derivative along the Voronoi edges in Fig. 6 (as opposed to the apparent discontinuous derivatives across Voronoi edges in Fig. 5) can be explained as follows. Notice that for all inscribing circle centers located in the interior of a Voronoi edge, there are only two points of contact from the input set and these two points remain invariant as long as the centers lie in the interior of that Voronoi edge. This implies that the 'active' constraints in Eq. (1) (that is, those constraints that turn to equality constraints) remain invariant during the entire traversal of the circle center in the interior of that Voronoi edge, thus contributing to the smoothness of the objective function and its directional derivatives along that edge.

Another interesting observation about the objective function along a Voronoi edge is that it achieves a minimum at an endpoint of that edge or at only one point in the interior of that edge. (It may also achieve a maximum, but that is not germane to the minimization search.) This property can be termed a 'quasi-min unimodality,' and can be exploited in employing the `fminbnd` function.

Since the gradient is also continuous along a Voronoi edge, a faster minimization method that employs the gradient could have been used. However, `fminbnd`, which requires only the objective function evaluation in its golden section search, is used in the code for simplicity and it seems to be sufficient for finding the minimum along each edge. Also, the 'quasi-min unimodality' mentioned above ensures that `fminbnd` will find the global minimum along each Voronoi edge. The global minimum among all Voronoi edges is extracted in line 31 of Table 2. In this simple implementation in line 31, only one of the minimum value is picked, even if there were multiple minima of the same value.

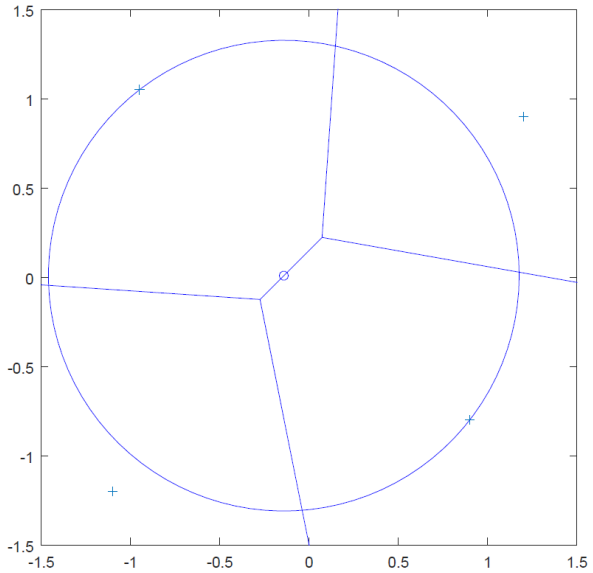


Figure 7. Results from A_CL2IC for Example 1.

The output circle from executing the code for the input from Example 1 is shown in Fig. 7, along with the Voronoi diagram of the input points. The center of the circle, shown as a small circle, lies in the interior of a small Voronoi edge, and the circle contacts two of the input points. For input points that may come from only an arc of a circle, consider the following example problem, again with only four points in a plane:

Example E2. $\boldsymbol{p_1}$: $(0.1, -0.9)$, $\boldsymbol{p_2}$: $(0.7, -0.75)$, $\boldsymbol{p_3}$: $(0.95, 0.1)$, and $\boldsymbol{p_4}$: $(0.65, 0.72)$.

Figure 8 shows the output circle from executing the code in Table 2 for the input points from Example 2, along with the Voronoi diagram of the input points. In Fig. 7, the center of the inscribing circle falls inside the convex hull of the input points. In Fig. 8, on the other hand, the center of the inscribing circle falls outside the convex hull of the input points.

The code for A_CL2IC given in Table 2 can be modified to obtain the code for A_CL2CC with only two changes:

1. In line 20, `voronoi` should be replaced by an appropriate call to a function that computes the furthest-neighbor Voronoi diagram. Such a code is available in `qhull` [14].
2. In line 49, `min(r)` should be replaced by `max(r)` so that the correct objective function for circumscribing circle is computed.

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

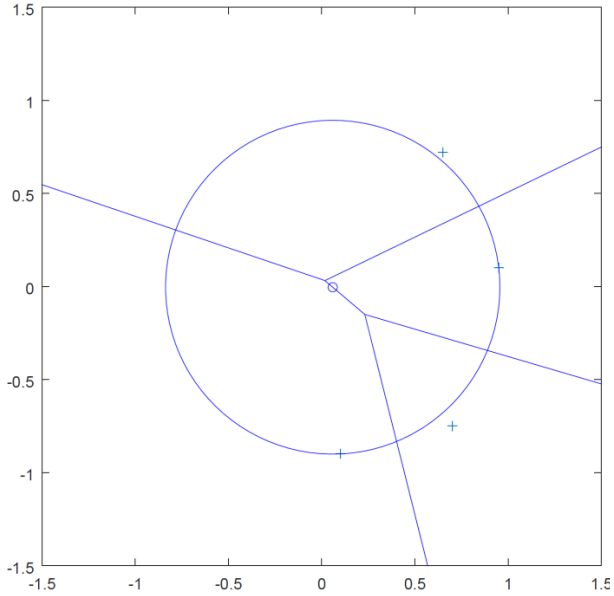


Figure 8. Results from A_CL2IC for Example 2.

Developing similar codes for A_CL2ISp and A_CL2CSp requires more work.

- For A_CL2ISp, GNU Octave has a `voronoin` function to compute the three-dimensional nearest-neighbor Voronoi diagram. For A_CL2CSp, a similar function to compute the three-dimensional furthest-neighbor Voronoi diagram can be found in `qhull` [14].
- In addition, for both A_CL2ISp and A_CL2CSp, `fminbnd` should be replaced by another minimizer for a smooth bivariate function over a bounded domain (in this case, a Voronoi face).

It is quite remarkable that a working code that implements an algorithm for CL2IC can be developed that is as compact as the one exhibited in Table 2. Even more remarkable are the power of heuristics to compute *all* the circles and spheres defined in Section 2, and the compactness of codes to implement them, as described in the next section.

4. Heuristics and their Implementations

The optimization problems posed in Eq. (2) can be attacked by heuristic methods that may be simpler than the algorithmic methods of Section 3. But these heuristics may not provide any theoretical guarantee about the global optimality of the solutions. Nevertheless, heuristics deserve attention because of their ease of implementation to solve a wide set of problems.

There are direct search methods, such as polytope (also known as simplex) methods, to find a minimum [16]. An outline of heuristics to attack the optimization problems of Eq. (2) is provided as Heuristic H_CL2x in Table 3.

Table 3. Heuristic for circles and spheres.

Heuristic H_CL2x (x stands for IC, CC, ISp, or CSp)
Input: Set of n points, either in a plane or in space.
Output: Center of optimum circle/sphere, its radius, and the minimized objective function.

1. Find a good starting solution set.
   a. For circles and spheres, this can be accomplished automatically and algorithmically using parabolic projection, as described in Section 4.1.
   b. A compact implementation of this algorithm is also described in Section 4.1.
2. Employ a derivative-free, direct search method to descend to a minimum. One such method is the following.
   a. Create a starting simplex around the starting solution set and move it by reflection, expansion, contraction, and scaling towards a minimum. Stop the search when it reaches a threshold.
   b. An implementation that employs Nedler-Mead method to accomplish this direct search is described in Section 4.2

Implementations of the two steps outlined in Heuristic H_CL2x are now described in Sections 4.1 and 4.2.

4.1 Finding good starting circles and spheres

Good starting circles and spheres as initial approximations are critical to heuristics that then find a local minimum. Luckily, there exists a mathematically elegant method that finds starting circles and spheres automatically and algorithmically by combining existing ideas in literature [2, 13]. A general algorithm to find an approximate sphere in m-dimensional space is described in Table 4 as Algorithm AppSphm, which can be specialized by setting $m = 2$ to find an approximate starting circle in a plane, and $m = 3$ to find an approximate starting sphere in space.

A visual interpretation of the Algorithm AppSphm, and a proof of its correctness, can be given using Fig. 9 for a simple case with $m = 1$, and generalizing the results to any m. Figure 9 shows a unit parabola U_2 in $\mathbb{R}^2$, and a point $x_1 = c_1 = 5$ in $\mathbb{R}^1$ that is projected up to the parabola in $\mathbb{R}^2$ as the point $(c_1, c_1^2) = (5, 25)$. Also shown in Fig. 9 is a line L_{1c} tangential to the parabola at the point (c_1, c_1^2), and another line L_1 that is parallel to L_{1c} but shifted up by a positive scalar amount equal to $r^2 = 9$. These entities can be described mathematically as:

$$U_2: x_2 = x_1^2 \tag{3}$$

$$L_{1c}: x_2 = 2c_1x_1 - c_1^2 \tag{4}$$

$$L_1: x_2 = 2c_1x_1 - c_1^2 + r^2. \tag{5}$$

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

The intersection of the unit parabola U_2 and the line L_1 can then be seen to be

$$U_2 \cap L_1 : (x_1 - c_1)^2 = r^2 \tag{6}$$

which can be interpreted as the equation of a one-dimensional sphere in $\mathbb{R}^1$ with radius r and centered at c_1.

Table 4. Algorithm for approximate circles and spheres.

Algorithm AppSphm

Input: Set of n points in $\mathbb{R}^m$. Each point $\boldsymbol{p_i}$ has coordinates $(x_{i,1}, x_{i,2}, \dots, x_{i,m})$ for $i = 1, n$. Let $n > m$.

Output: Center coordinates $(c_1, c_2, \dots, c_m)$ and radius r_c of a sphere in $\mathbb{R}^m$.

1. Project points in $\mathbb{R}^m$ to a unit paraboloid in $\mathbb{R}^{m+1}$. This is accomplished by setting $x_{m+1} = x_1^2 + x_2^2 + \cdots + x_m^2$ for each point.
2. Fit a least-squares hyperplane to these points in $\mathbb{R}^{m+1}$. This is accomplished by solving a set of over-constrained linear equations
$$\begin{bmatrix} 2x_{1,1} & \dots & 2x_{1,m} & 1 \\ 2x_{2,1} & \dots & 2x_{2,m} & 1 \\ \dots & \dots & \dots & \dots \\ 2x_{n,1} & \dots & 2x_{n,m} & 1 \end{bmatrix} \begin{Bmatrix} c_1 \\ \dots \\ c_m \\ c_{m+1} \end{Bmatrix} = \begin{Bmatrix} x_{1,1}^2 + \cdots + x_{1,m}^2 \\ \dots \\ x_{n,1}^2 + \cdots + x_{n,m}^2 \end{Bmatrix}$$
using the least-squares method.
3. Find the intersection of the unit paraboloid and the hyperplane. It is an m-dimensional ellipsoid in $\mathbb{R}^{m+1}$.
4. Project that ellipsoid back to $\mathbb{R}^m$ to obtain a sphere in $\mathbb{R}^m$, and return its center coordinates $(c_1, c_2, \dots, c_m)$ and radius $r_c = \sqrt{c_1^2 + \cdots + c_m^2 + c_{m+1}}$.

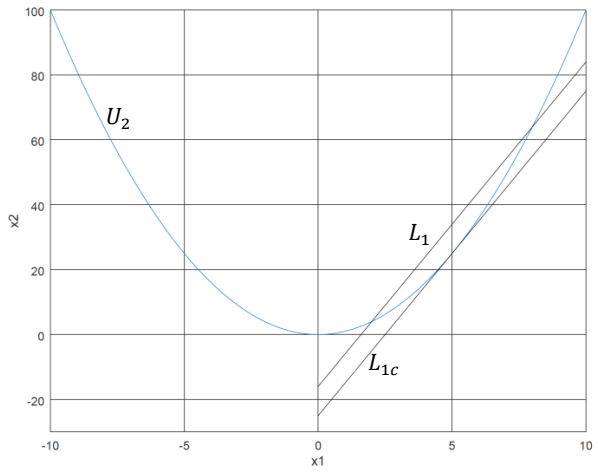


Figure 9. Parabola and parallel lines.

The Eqs. (3-6) can be generalized to any dimension m by considering a unit paraboloid U_{m+1} in $\mathbb{R}^{m+1}$, and a point $(c_1, c_2, \dots, c_m) \in \mathbb{R}^m$ that is projected up to the paraboloid in $\mathbb{R}^{m+1}$ as the point $\left(c_1, c_2, \dots, c_m, \sum_{j=1}^m c_j^2\right)$. Also consider a hyperplane P_{mc} tangential to the paraboloid at the point $\left(c_1, c_2, \dots, c_m, \sum_{j=1}^m c_j^2\right)$, and another hyperplane P_m that is parallel to P_{mc} but shifted up by a scalar amount equal to r^2. These entities can be described mathematically as:

$$U_{m+1} : x_{m+1} = \sum_{j=1}^m x_j^2 \tag{7}$$

$$P_{mc} : x_{m+1} = 2\sum_{j=1}^m c_j x_j - \sum_{j=1}^m c_j^2 \tag{8}$$

$$P_m : x_{m+1} = 2\sum_{j=1}^m c_j x_j - \sum_{j=1}^m c_j^2 + r^2 \tag{9}$$

The intersection of the unit paraboloid U_{m+1} and the hyperplane P_m can then be seen to be

$$U_{m+1} \cap P_m : \sum_{j=1}^m \left(x_j - c_j\right)^2 = r^2 \tag{10}$$

which can be interpreted as the equation of a m-dimensional sphere in $\mathbb{R}^m$ with radius r and centered at $(c_1, c_2, \dots, c_m)$. When $m = 2$ it is a circle in a plane, and when $m = 3$ it is a sphere in three-dimensional space.

With these preliminary results, the proof of correctness of Algorithm AppSphm can be obtained by considering fitting a hyperplane

$$x_{m+1} = 2\sum_{j=1}^m x_j c_j + c_{m+1} \tag{11}$$

to a set of n points in $\mathbb{R}^{m+1}$. These points are obtained in Step 1 of AppSphm by the mapping

$$(x_{i,1}, x_{i,2}, \dots, x_{i,m}) \rightarrow \left(x_{i,1}, x_{i,2}, \dots, x_{i,m}, \sum_{j=1}^m x_{i,j}^2\right) \tag{12}$$

for all i. When these point coordinates are applied to Eq. (11), they result in a set of over-determined equations (because it is assumed that $n > m$, and often $n \gg m$) shown in Step 2 of AppSphm. In Step 3, these over-determined equations are solved using the usual least-squares methods to obtain the coefficients $c_j,\ j = 1, \dots, m+1$. Then, Eq. (10) implies that the center of the desired sphere has the coordinates $(c_1, c_2, \dots, c_m)$, and a comparison of Eq. (9) and Eq. (11) indicates that the radius r of the sphere can be obtained from $c_{m+1} = -\sum_{j=1}^m c_j^2 + r^2$. This proves the correctness of the last step (Step 4) of AppSphm, and thus of the whole algorithm.

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

Table 5. Code for AppCir

filename: AppCir.m

```
function [c,rc]=AppCir(M)
  %
  % Computes an approximate circle in a plane based on (1) parabolic projection
  % of 2D points to 3D, (2) fitting an ordinary least-squares plane in 3D, and
  % (3) projecting the intersecting ellipse to the 2D plane.
  % Input:
  %   M: an nx2 matrix of x and y coordinates of n points
  % Output:
  %   c: center of the circle; a 1x2 array of x and y coordinates
  %   rc: radius of the circle; a scalar
  %
  x = [2*M  ones(size(M)(1),1)]\(M(:,1).^2 + M(:,2).^2); % Does the magic!
  c = [x(1) x(2)]; % Center of the circle
  rc = sqrt(x(1)^2 + x(2)^2 + x(3)); % Radius of the circle
end
```

Table 6. Code for H_CL2IC

filename: H_CL2IC.m

```
function [c,rc,f]=H_CL2IC(M)
%
% Implementation of a heuristic
%   to fit a constrained (inscribed) least-squares circle in a plane
% Input:
%   M: an nx2 matrix of x and y coordinates of n points
% Output:
%   c:  center of the circle; a 1x2 array of x and y coordinates
%   rc: radius of the circle; a scalar
%   f:  minimized objective function; a scalar
%
% Needs:
%   AppCir: an external function to get an approximate circle
%   fminsearch: a built-in function to search for minimum
%   CL2IC_ObjFun: a local function needed by fminsearch
%
  global P % Make the coordinate matrix a global variable in this file
  P = M; % Copy M to P
  [x0, apprad] = AppCir(P); % Get the approximate circle to start the search
  [c, f] = fminsearch(@CL2IC_ObjFun, x0); % Conduct the heuristic search
%   Input:
%     @CL2IC_ObjFun: handle for a local objective function
%     x0: a 1x2 array of starting circle center coordinates
%   Output:
%     c: a 1x2 array of circle center coordinates
%     f: minimized objective function; a scalar
  rc = min(sqrt((P(:,1)-c(1)).^2 + (P(:,2)-c(2)).^2)); % Radius of the circle
end
%
  function f=CL2IC_ObjFun(x)
%   Objective function needed by fminsearch
%   Input:
%     x: starting approximation for center and radius
%     center is a 1x2 array of x and y coordinates
%     radius is a scalar
%   Output:
%     f: Objective function = square root of the mean of the squares (RMS)
%        of the deviations of the points from the circle; a scalar
%
    global P % a global variable to get the coordinate matrix
    r = sqrt((P(:,1)-x(1)).^2 + (P(:,2)-x(2)).^2); % radius vector
    f = (norm(r-min(r)))/sqrt(size(r)(1)); % RMS of deviations from the circle
  end
```


Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

Table 5 exhibits a compact GNU Octave code for the AppCir algorithm, which specializes Algorithm AppSphm for $m = 2$. It is remarkable that the entire computation is executed in a single line (line 12). A similar code that specializes Algorithm AppSphm for $m = 3$ is given in the Annex Table A2.

4.2 Direct search to find a local minimum

The fact that the objective functions of Eq. (2) have discontinuous gradient vectors, and that the minimum always occurs at these discontinuities (that is, on Voronoi diagrams) limits the choice of generic optimization methods. For problems of these types, one may resort to a direct search method that does not depend on derivatives. One such derivative-free direct search can be carried out by the Nedler-Mead method [16], which has been well documented and implemented in publicly-available software.

A code that implements the Heuristic H_CL2IC in GNU Octave is exhibited in Table 6. It uses the Nedler-Mead method by invoking a built-in function `fminsearch` in line 20. Similar code for heuristic H_CL2CC for constrained least-squares fitting of circumscribing circles is exhibited in Annex Table A1. The only difference between the codes in Table 6 and Table A1 is in line 42, where the function `min(r)` is replaced by `max(r)` in evaluating the objective function.

Figures 10 and 11 show the results of executing the code H_CL2IC for Example 1 and Example 2, respectively. Contour plots of the objective functions are superposed in these figures to illustrate how the centers (marked as 'o') are found by the search heuristics at the 'bottom of the valley.'

It is instructive to compare Figs. 7 and 10 for Example 1, and Figs. 8 and 11 for Example 2. The results of the algorithmic and heuristic codes look strikingly identical. Table 7 summarizes and compares the numerical results from algorithmic and heuristic computations of center coordinates (c), radii (r), and objective functions (f) for inscribing circles; also presented are the center and radius values for approximate starting circles used in search heuristics. The numerical results differ within the default tolerances set in the GNU Octave built-in functions `fminbnd` and `fminsearch`.

Table 7. Comparison of numerical results**.**

	Example 1	Example 2
A_CL2IC	$c = (-0.1417377,$ $0.0082623)$ $r = 1.3185$ f =0.18410	$c = (0.0592824,$ $-0.0041400)$ $r = 0.89678$ $f = 0.047181$
H_CL2IC	$c = (-0.1416265,$ $0.0083757)$ r =1.3185 f =0.18410	$c = (0.0596950,$ $-0.0044904)$ $r = 0.89642$ $f = 0.047181$
AppCir	$c = (-0.1061450,$ $0.0049736)$ $r = 1.4501$	$c = (0.107328,$ $-0.023512)$ $r = 0.89722$

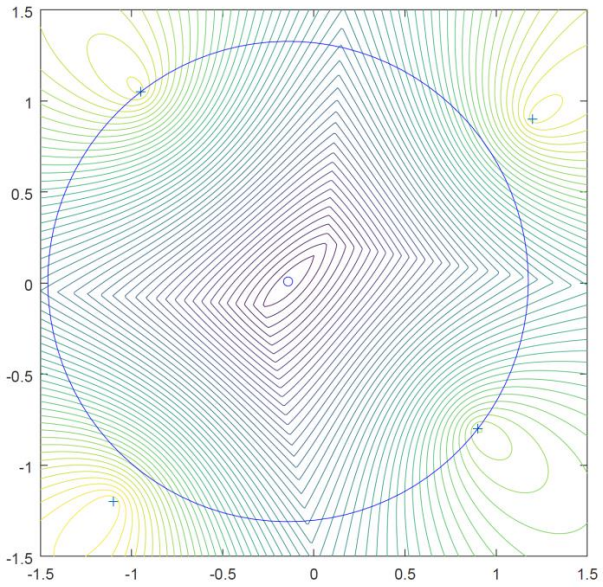


Figure 10. Results from H_CL2IC for Example 1.

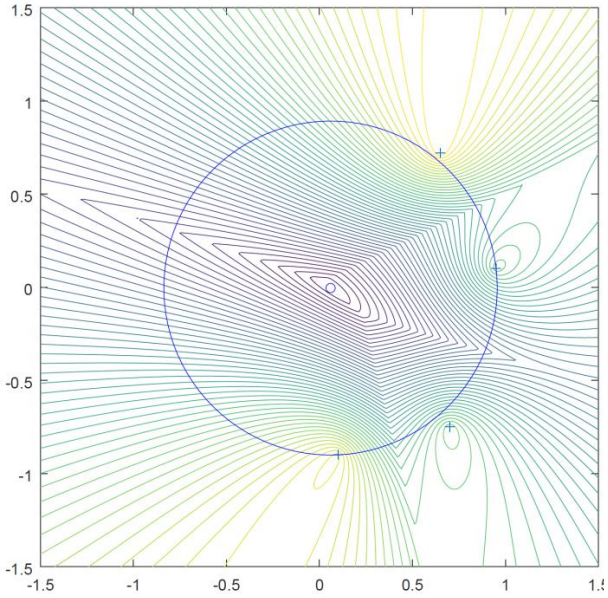


Figure 11. Results from H_CL2IC for Example 2.

To complete the picture, Figs. 12 and 13 illustrate the results of executing the code H_CL2CC for Examples 1 and 2. Again, it can be observed that the search heuristic solution is centered at the 'bottom of the valley.' One way to interpret the working of the Nedler-Mead method using Figs.10 to 13 is to imagine a

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

water droplet deposited at the center of the approximate starting circle. This droplet, taking the form of a changing simplex, then trickles down by gravity to the bottom of the valley. Tables A3 and A4 in the Annex exhibit the heuristics H_CL2ISp and H_CL2CSp, respectively, to compute the constrained least-squares fitting of inscribing and circumscribing spheres.

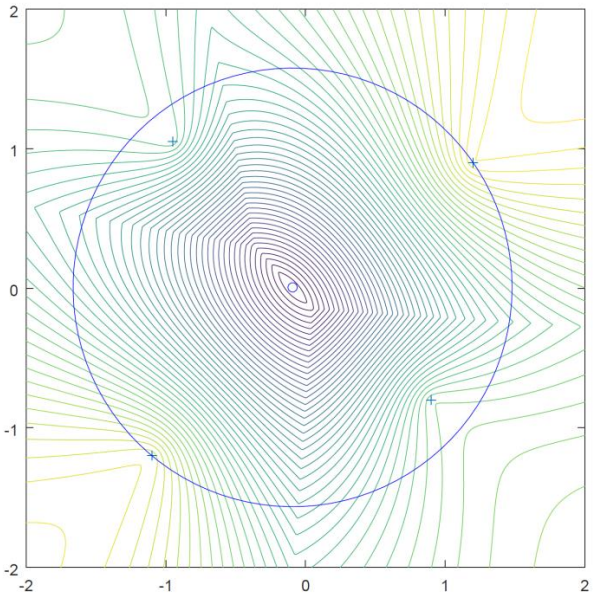


Figure 12. Results from H_CL2CC for Example 1.

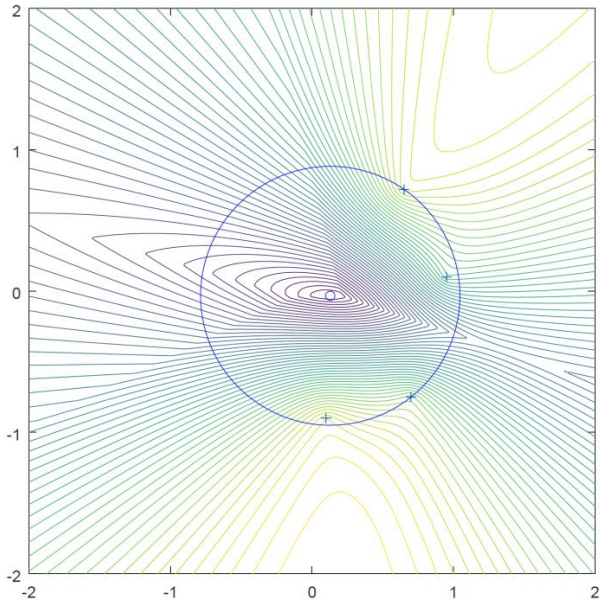


Figure 13. Results from H_CL2CC for Example 2.

5. Opportunities for Improvements and Extensions

The implementations of algorithms and heuristics presented in this paper can be improved in many ways. Parallel processing using GPUs can obviously accelerate the computation of Voronoi diagrams, and search for minima over Voronoi edges and Voronoi faces in the algorithms. Any number of direct search methods may be improved and deployed to find the minima in the heuristics.

An exciting opportunity lies in the extension of the heuristics to compute the constrained least-squares fitting of other geometric elements such as cylinders, cones, tori, and free-form surfaces. The ease of implementation of the heuristics for circles and spheres reported in this paper gives some encouragement to such extensions. However, these extensions should be subjected to careful analysis and testing before they can be adopted for serious use in industry.

6. Summary and Concluding Remarks

This paper addressed the practical issues in implementing constrained least-squares fitting of circles and spheres. These problems have acquired some urgency due to the impending adoption of the constrained least-squares criterion as the common, default definition for datums in ASME and ISO standards for geometric dimensioning and tolerancing.

In addition to presenting algorithms and heuristics, the paper also provided representative codes written in freely available GNU Octave language to encourage software testing and adoption by industry. As indicated in Section 5, exciting opportunities exist to extend the heuristics for constrained least-squares fitting to other important geometric elements covered by standards. This promises to be an area for further fruitful research.

Acknowledgment and Disclaimer

The authors thank ISO and ASME standards experts whose advice and suggestions were invaluable in initiating and sustaining this research investigation. Any mention of commercial products or systems in this article is for information only; it does not imply recommendation or endorsement by NIST. The software codes presented in this paper come with absolutely no warranty.

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

References

[1] Shakarji, C.M. and Srinivasan, V., Toward a new mathematical definition of datums in standards to support advanced manufacturing, MSEC2018-6305, Proceedings of the ASME 2018 Manufacturing Science and Engineering Conference, College Station, Texas, June 18-22, 2018.

[2] Forbes, A.B., Least-squares best-fit geometric elements, NPL Report DITC 140/89, Revised edition Feb. 1991, National Physical Laboratory, U.K., 1991.

[3] Shakarji, C.M., Least-squares fitting algorithms of the NIST algorithm testing system, Journal of Research of the National Institute of Standards and Technology, Vol. 30, pp. 663-641, Dec. 1998.

[4] Srinivasan, V., Shakarji, C.M. and Morse, E.P., On the enduring appeal of least-squares fitting in computational coordinate metrology, ASME Journal of Computing and Information Science in Engineering, Vol. 12, March 2012.

[5] Shakarji, C.M. and Srinivasan, V., Theory and algorithms for weighted total least-squares fitting of lines, planes, and parallel planes to support tolerancing standards, ASME Journal of Computing and Information Science in Engineering, 13(3), 2013.

[6] Coons, S.A., Constrained least-squares, Computers & Graphics, Vol. 3, No. 1, pp. 43-47, 1978.

[7] Golub, G.H., and Matt, U.V., Quadratically constrained least squares and quadratic problems, Numerische Mathematik, Vol. 59, No. 1, pp. 561-580, 1991.

[8] Peng, J.J., and Liao, A.P., Algorithm for inequality-constrained least squares problems, Computational and Applied Mathematics, Vol. 36, No. 1, pp. 249-258, 2017.

[9] Shakarji, C.M. and Srinivasan, V., A constrained L_2 based algorithm for standardized planar datum establishment, ASME IMECE2015-50654, Proceedings of the ASME 2015 International Mechanical Engineering Congress and Exposition, Houston, TX, Nov. 13-19, 2015.

[10] Shakarji, C.M. and Srinivasan, V., Theory and algorithm for planar datum establishment using constrained total least-squares, 14th CIRP Conference on Computer Aided Tolerancing, Gothenburg, Sweden, 2016.

[11] Shakarji, C.M. and Srinivasan, V., Convexity and optimality conditions for constrained least-squares fitting of planes and parallel planes to establish datums, IMECE2017-70899, Proceedings of the ASME 2017 International Mechanical Engineering Congress and Exposition, Tampa, FL, Nov. 3-9, 2017.

[12] Shakarji, C.M. and Srinivasan, V., Optimality conditions for constrained least-squares fitting of circles, cylinders, and spheres to establish datums, ASME Journal of Computing and Information Science in Engineering, 2018. (Also DETC2017-67143, Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Cleveland, OH, Aug. 6-9, 2017.)

[13] O'Rourke, J., Computational Geometry in C, 2nd Edition, Cambridge University Press, 1998.

[14] www.qhull.org

[15] Eaton, J.W., Bateman, D., Hauberg, S., and Wehbring, R., GNU Octave version 4.2.0 manual: a high-level interactive language for numerical computations, 2016. url: www.gnu.org/software/octave/doc/interpreter/

[16] Gill, P.E., Murray, W., and Wright, M.H., Practical Optimization, Emerald Group Publishing Ltd., 1982.

Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

Annex

Table A1. Code for H_CL2CC

filename: H_CL2CC.m

```
function [c,rc,f]=H_CL2CC(M)
%
% Implementation of a heuristic
%   to fit a constrained (circumscribed) least-squares circle in a plane
% Input:
%   M: an nx2 matrix of x and y coordinates of n points
% Output:
%   c:  center of the circle; a 1x2 array of x and y coordinates
%   rc: radius of the circle; a scalar
%   f:  minimized objective function; a scalar
%
% Needs:
%   AppCir: an external function to get an approximate circle
%   fminsearch: a built-in function to search for minimum
%   CL2CC_ObjFun: a local function needed by fminsearch
%
  global P % Make the coordinate matrix a global variable in this file
  P = M; % Copy M to P
  [x0, apprad] = AppCir(P); % Get the approximate circle to start the search
  [c, f] = fminsearch(@CL2CC_ObjFun, x0); % Conduct the heuristic search
%   Input:
%     @CL2CC_ObjFun: handle for a local objective function
%     x0: a 1x2 array of starting circle center coordinates
%   Output:
%     c: a 1x2 array of circle center coordinates
%     f: minimized objective function; a scalar
  rc = max(sqrt((P(:,1)-c(1)).^2 + (P(:,2)-c(2)).^2)); % Radius of the circle
end
%
  function f=CL2CC_ObjFun(x)
%   Objective function needed by fminsearch
%   Input:
%     x: starting approximation for center and radius
%     center is a 1x2 array of x and y coordinates
%     radius is a scalar
%   Output:
%     f: Objective function = square root of the mean of the squares (RMS)
%        of the deviations of the points from the circle; a scalar
%
    global P % a global variable to get the coordinate matrix
    r = sqrt((P(:,1)-x(1)).^2 + (P(:,2)-x(2)).^2); % radius vector
    f = (norm(r-max(r)))/sqrt(size(r)(1)); % RMS of deviations from the circle
  end
```

Table A2. Code for AppSph

filename: AppSph.m

```
function [c,r]=AppSph(M)
  %
  % Computes an approximate sphere in space based on (1) parabolic projection
  % of 3D points to 4D, (2) fitting an ordinary least-squares hyper-plane in 4D,
  % and(3) projecting the intersecting ellipsoid to the 3D space.
  % Input:
  %   M: an nx3 matrix of x,y and z coordinates of n points
  % Output:
  %   c: center of the sphere; a 1x3 array of x,y and z coordinates
  %   r: radius of the sphere; a scalar
  %
   x = [2*M  ones(size(M)(1),1)]\(M(:,1).^2 + M(:,2).^2 + M(:,3).^2); % Does the magic!
   c = [x(1) x(2) x(3)]; % Center of the sphere
   r = sqrt(x(1)^2 + x(2)^2 + x(3)^2 + x(4)); % Radius of the sphere
end
```


Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

Table A3. Code for H_CL2ISp

filename: H_CL2ISp.m

```
function [c,rc,f]=H_CL2ISp(M)
%
% Implementation of a heuristic
%   to fit a constrained (inscribed) least-squares sphere in space
% Input:
%   M: an nx3 matrix of x,y and z coordinates of n points
% Output:
%   c:  center of the sphere; a 1x3 array of x,y and z coordinates
%   rc: radius of the sphere; a scalar
%   f:  minimized objective function; a scalar
%
% Needs:
%   AppSph: an external function to get an approximate circle
%   fminsearch: a built-in function to search for minimum
%   CL2ISp_ObjFun: a local function needed by fminsearch
%
  global P % Make the coordinate matrix a global variable in this file
  P = M; % Copy M to P
  [x0, apprad] = AppSph(P); % Get the approximate sphere to start the search
  [c, f] = fminsearch(@CL2ISp_ObjFun, x0); % Conduct the heuristic search
%   Input:
%     @CL2ISp_ObjFun: handle for a local objective function
%     x0: a 1x3 array of starting sphere center coordinates
%   Output:
%     c: a 1x3 array of sphere center coordinates
%     f: minimized objective function; a scalar
  rc = min(sqrt((P(:,1)-c(1)).^2 + (P(:,2)-c(2)).^2 + (P(:,3)-c(3)).^2)); % Radius of the sphere
end
%
  function f=CL2ISp_ObjFun(x)
%   Objective function needed by fminsearch
%   Input:
%     x: starting approximation for center and radius
%     center is a 1x3 array of x,y and z coordinates
%     radius is a scalar
%   Output:
%     f: Objective function = square root of the mean of the squares (RMS)
%        of the deviations of the points from the sphere; a scalar
%
    global P % a global variable to get the coordinate matrix
    r = sqrt((P(:,1)-x(1)).^2 + (P(:,2)-x(2)).^2 + (P(:,3)-x(3)).^2); % radius vector
    f = (norm(r-min(r)))/sqrt(size(r)(1)); % RMS of deviations from the sphere
  end
```


Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.

Table A4. Code for H_CL2CSp

filename: H_CL2CSp.m

```
function [c,rc,f]=H_CL2CSp(M)
%
% Implementation of a heuristic
%   to fit a constrained (circumscribed) least-squares sphere in space
% Input:
%   M: an nx3 matrix of x,y and z coordinates of n points
% Output:
%   c:  center of the sphere; a 1x3 array of x,y and z coordinates
%   rc: radius of the sphere; a scalar
%   f:  minimized objective function; a scalar
%
% Needs:
%   AppSph: an external function to get an approximate circle
%   fminsearch: a built-in function to search for minimum
%   CL2CSp_ObjFun: a local function needed by fminsearch
%
  global P % Make the coordinate matrix a global variable in this file
  P = M; % Copy M to P
  [x0, apprad] = AppSph(P); % Get the approximate sphere to start the search
  [c, f] = fminsearch(@CL2CSp_ObjFun, x0); % Conduct the heuristic search
%   Input:
%     @CL2CSp_ObjFun: handle for a local objective function
%     x0: a 1x3 array of starting sphere center coordinates
%   Output
%     c: a 1x3 array of sphere center coordinates
%     f: minimized objective function; a scalar
  rc = max(sqrt((P(:,1)-c(1)).^2 + (P(:,2)-c(2)).^2 + (P(:,3)-c(3)).^2)); % Radius of the sphere
end
%
  function f=CL2CSp_ObjFun(x)
%   Objective function needed by fminsearch
%   Input:
%     x: starting approximation for center and radius
%     center is a 1x3 array of x,y and z coordinates
%     radius is a scalar
%   Output:
%     f: Objective function = square root of the mean of the squares (RMS)
%        of the deviations of the points from the sphere; a scalar
%
    global P % a global variable to get the coordinate matrix
    r = sqrt((P(:,1)-x(1)).^2 + (P(:,2)-x(2)).^2 + (P(:,3)-x(3)).^2); % radius vector
    f = (norm(r-max(r)))/sqrt(size(r)(1)); % RMS of deviations from the sphere
  end
```


Shakarji, Craig; Srinivasan, Vijay.
"On Algorithms and Heuristics for Constrained Least-Squares Fitting of Circles and Spheres to Support Standards."
Paper presented at ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Quebec City, QC, Canada. August 26, 2018 - August 29, 2018.