Newsjournal of the Society for Industrial and Applied Mathematics sinews.siam.org

# siam news

Volume 54/ Issue 7
September 2021

# The Loopy Fluid Dynamics of Bird Lungs

By Anand U. Oza, Christina Frederick, and Leif Ristroph

If asked why birds are such efficient aerodynamic machines, one may initially point to their streamlined bodies and finely contoured wings that seem tailored to manipulate the wind. But the aerodynamics behind birds' mastery of flight is as much about the intricate structures and airflows *inside* as well as outside. Investigations into the nonlinear dynamics of flow networks with unusual topologies are breathing fresh air into a century-old mystery about bird lungs [7].

Biologists have long appreciated birds' ability to breathe highly efficiently — a trait that is presumably connected with the extreme energetic demands of flight [5]. A classic experiment illustrates this point by stark comparison: in an enclosed chamber where oxygen has been depleted to such a level that a mouse is comatose and moribund, a similarly-sized sparrow playfully flitters about [10]. These remarkable respiratory abilities allow birds to migrate across oceans and continents and fly in thin air at elevations as high as Mount Everest [9]. It is thus with good reason that anatomists have carefully mapped out the structure of avian lungs and physiologists have charted the airflow patterns and currents [2].

The airflow in the lungs of most animals (including humans) oscillates with inhalation and exhalation. However, one-way or directed flows persist throughout the bird respiration cycle due to a unique topology that—unlike our own branched bronchial network—contains looped airways (see Figure 1a). The one-way flows arise as circulation around these loops, meaning that the time average of the fluid flux over a breathing cycle has a nonzero value. Bird lungs therefore exhibit a form of alternating current (AC) to direct current (DC) rectification, wherein the oscillatory AC airflow that is associated with inhalation and exhalation is converted to a directed DC flow around the loops. While researchers are still parsing through the details [6], it seems that this one-way flow of air allows for efficient gas exchange with the blood, which itself flows directionally within the circulatory system because of the one-way action of heart valves.

Lung valves, however, seem to be nothing but science fiction. In the 100 years since scientists first documented the phenomenon of one-way airflow, exhaustive searches for valvular structures by anatomists have come up empty [4]. Consequently, researchers have proposed many hypotheses that appeal to various structural features in the lung that may appropriately guide the flows, even without the opening and closing movements of conventional valves [1]. However, because anatomical details differ across species yet one-way flows are pervasive, we felt that a more general mechanism lurks in the fluid dynamics of loopy networks.

Rather than use a detailed model of a bird lung that captures all anatomical complexities, our recent study considers a "spherical bird" approximation that isolates a single loop in the network (see Figure 1b). We then transformed the system into the closed circuit in Figure 1c — an experimental and computational convenience for the study of internal flows. The network contains a fluid that is driven through the action of an oscillating piston, which plays the role of inhalation and exhalation in bird lungs. This system allows

See **Bird Lungs** on page 4

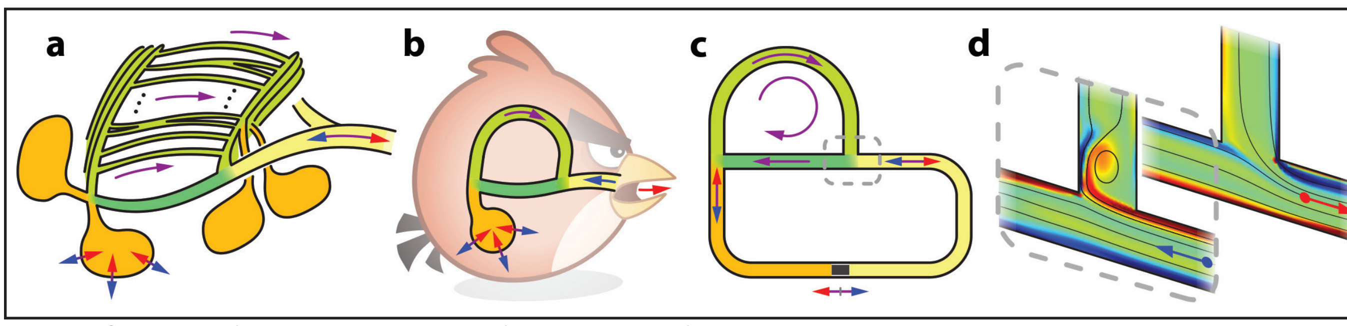



**Figure 1.** *Oscillatory airflows during breathing transform into one-way flows around loops in the bird lung.* **1a.** *Unlike the branched network of mammalian lungs, avian lungs have loops and reconnections.* **1b.** *A "spherical bird" approximation to the lung network that involves one loop (green) and oscillatory forcing from a single air sac (orange).* **1c.** *A closed-circuit system that is driven by a reciprocating piston.* **1d.** *Simulations show that nonlinear effects at network junctions provide the valving action that directs flows around the loop. Figure courtesy of the authors.*

# Digital Twins: Where Data, Mathematics, Models, and Decisions Collide

By Michael G. Kapteyn and Karen E. Willcox

Regardless of whether one is entering the interdisciplinary field of computational science and engineering with a background in mathematics, statistics, computer science, engineering, or the sciences—or maybe a little bit of each—it is such an exciting time to be a computational scientist. The field is in the midst of a tremendous convergence of *technologies* that generate unprecedented system data and enable automation, *algorithms* that let users process massive amounts of data and run predictive simulations that drive key decisions, and the *computing power* that makes these algorithms feasible at scale for complex systems and in real-time or in situ settings.

The convergence of these three elements is a revolution that touches almost every area of science, engineering, and society. It affects all aspects of a complex system's lifecycle by advancing simulation and modeling capabilities in each phase of design; impacting the way we plan, monitor, and optimize manufacturing processes; and changing the way we operate systems.

## Digital Twins Integrate Data, Models, and Decisions

A digital twin is defined as "a set of virtual information constructs that mimics the structure, context, and behavior of an individual/unique physical asset, is dynamically updated with data from its physical twin throughout its lifecycle, and informs decisions that realize value" [1]. People often wonder how a digital twin differs from the modeling and simulation that researchers have been conducting for decades. The quoted definition highlights three key points of differentiation. First, a digital twin is personalized because it targets an individual or unique asset instead of being a generic computational model. This means that it must reflect asset-to-asset differences and variability. Second, a digital twin is a living model—not a static computational model—that evolves as the physical twin evolves. And third, a digital twin encapsulates an integrated end-to-end view of data, models, and decisions.

The community is beginning to see digital twins deployed for myriad engineering applications, including aircraft, spacecraft, buildings, bridges, engines, automobiles, wind turbines, and floating production storage and offloading units. As Figure 1 illustrates, there is also gathering momentum to develop digital twins in medicine and the geosciences. However, state-of-the-art digital twins are largely the result of custom implementations that require considerable deployment resources and a high level of expertise. Moving from the one-off digital twin to accessible robust digital twin implementations at scale requires rigorous and scalable mathematical foundations [4].

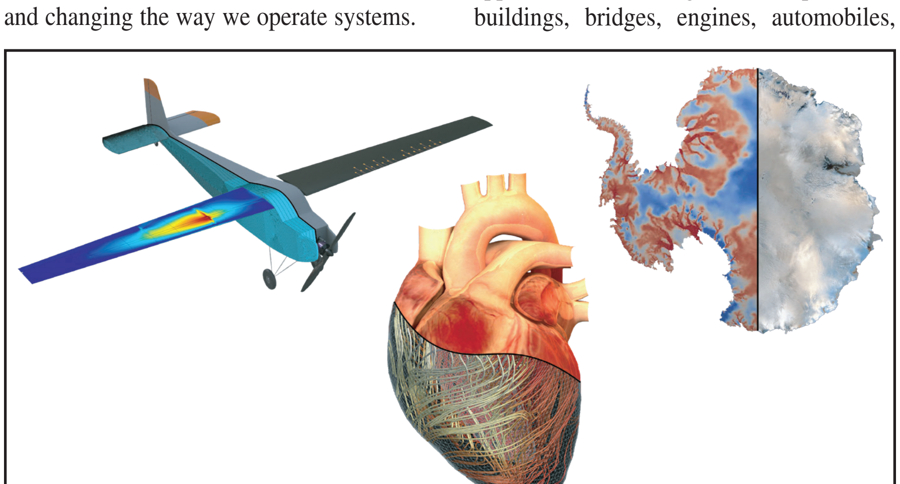


**Figure 1.** *Digital twins can enable asset-specific monitoring, analysis, prediction, and intervention in a range of applications, including engineering systems like aircraft and natural systems like biological organs and environmental systems. Aircraft image adapted from [3], heart model courtesy of Michael Sacks and Greg Foss, and Antarctica model courtesy of Omar Ghattas.*

## A Mathematical Foundation for Digital Twins

One example of a mathematical foundation for digital twins is a probabilistic graphical model [3]. Probabilistic graphical models provide a powerful mathematical abstraction for modeling complex systems in an insightful and intuitive way while also serving as a foundation for generalizable and scalable computational methods. For example, graphical models have seen wide application as a unified framework for modeling and inference in the field of robotics, where they enable tasks like perception, state estimation, and motion planning for a variety of robotic systems [2]. In the context of digital twins, a probabilistic graphical model enables data-driven asset monitoring; digital twin model updating; and model-based prediction, planning, and decision-making to all be formulated as probabilistic inference tasks. Furthermore, one can exploit the graphical model structure to develop principled and scalable algorithms that carry out these inference tasks.

To develop a probabilistic graphical model for digital twins, we first formulate a mathematical abstraction of the system, which is comprised of six key elements and defined in Figure 2 (on page 3). The partially observable physical asset state $S_t$ is reflected in the digital twin state $D_t$,

See **Digital Twins** on page 3

Nonprofit Org
U.S. Postage
PAID
Permit No 360
Bellmawr, NJ

siam
SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA

siam news
Volume 54/ Issue 7/ September 2021



# SIAM Names Tayler Fernandes Nuñez as Inaugural SIAM EDGE Fellow

*By Lina Sorg*

The EDGE Program,[1] which is administered by the Sylvia Bozeman and Rhonda Hughes EDGE Foundation, supports women who are pursuing careers in mathematical research and education. By sponsoring an annual conference, a summer session, mentoring activities, and research-related travel, the organization aims to increase the number of women in leadership roles throughout the mathematics community.

The EDGE Summer Program[2] is a rigorous month-long session for women who are entering Ph.D. programs in the mathematical sciences. The program, which takes place at a different university each year, is meant to prepare attendees for graduate-level research and qualification exams while simultaneously offering practical experience in a demanding academic environment. The 2021 EDGE Summer Program[3] took place in June at the University of Minnesota, Twin Cities.

This year, SIAM sponsored a SIAM EDGE Fellow for the first time. Tayler Fernandes Nuñez—the 2021 recipient—was selected by EDGE instructors and co-directors for her exceptional work throughout the summer program. *SIAM News* recently sat down with Tayler to discuss her undergraduate trajectory, experiences with and reflections of EDGE, and future plans as she begins graduate school at Cornell University.

***SIAM News*: How did you first hear about the EDGE Summer Program?**

**Tayler Fernandes Nuñez:** I found out about EDGE when I was a participant in the Pomona Research in Mathematics Experience[4]—a Research Experiences for Undergraduates (REU) program that focused on group theory and number theory—under the leadership of Edray Goins. It was the summer of 2019 and the EDGE Summer Program also took place at Pomona College that year, so I met some of the EDGE attendees. They all encouraged the REU students to apply. I pocketed the idea and waited until I finished the Postbaccalaureate Program at Smith College,[5] and I'm really happy that I kept it in the back of my head.

[1] https://www.edgeforwomen.org
[2] https://www.edgeforwomen.org/summer-session
[3] https://www.edgeforwomen.org/edge-2021-participants
[4] https://pages.pomona.edu/~ehga2017/prime.html


*Tayler Fernandes Nuñez, who is currently pursuing her Ph.D. at Cornell University, is the inaugural SIAM EDGE Fellow. Photo courtesy of the EDGE Foundation.*

**SN: Tell us about your undergraduate experiences. What kind of path led you to Edray Goins' REU program, Smith College, and ultimately to EDGE?**

**TFN:** I actually started at Northeastern University as a psychology major, but I found that I missed math. So I decided to take a logic course and I loved it. I knew just from the course that math is more artistic than it appears in the beginning stages, and I ended up switching my major to mathematics after my sophomore year. I attended a lot of conferences my senior year at Northeastern to showcase the research I did with Dr. Goins, and that's how I learned about Smith's Postbaccalaureate Program — a certificate program that involves a year of additional classes to help students prepare for graduate school. I earned a B.S. in mathematics from Northeastern and applied for Smith's program because I hadn't taken many advanced electives in mathematics as an undergraduate. I then applied for the 2021 EDGE Summer Program while earning my Smith certificate.

[5] https://www.smith.edu/academics/graduate/mathematics-postbaccalaureate


*Students, faculty, mentors, administrative staff, and EDGE alumnae celebrate with a banquet during the 2021 EDGE Summer Program, which took place in June at the University of Minnesota, Twin Cities. Photo courtesy of the EDGE Foundation.*

**SN: What was the structure of this year's EDGE Summer Program?**

**TFN:** It was a four-week sequence of four classes. The first two weeks focused on real analysis and algebra, and the last two weeks emphasized machine learning and measure theory. We had six lessons per class and classes were about 1.5 hours. Our mornings consisted of either classes or problem sessions. After lunch, we had colloquium speakers, MATLAB mini-courses, scheduled activities, or free time. Office hours with professors and mentors were in the evenings.

The problems that we worked on in class were difficult. There were between five and seven questions per problem set, six problem sets per class, and only two weeks for each class. The turnover was really fast.

**SN: Sounds like a pretty rigorous schedule. Was it difficult to manage such a heavy workload?**

**TFN:** I think that EDGE does a really good job of adding the stress factor that we need in order to make tough decisions about our learning processes. The pressure allowed me to be reflective of the way I address stress — it kind of shined a light into some of our strengths or weaknesses as individuals and also as learners.

EDGE also planned social events, so we had to figure out how to preserve our energy and recharge ourselves through other activities when we had work to do. In some ways, it was a smaller ecosystem of what we expect grad school to be like: lots of work, very intense, and difficult decisions on a day-to-day or week-to-week basis. And I think that was the purpose of the program. It's about the math, but it's more so about navigating who you are as a learner, figuring out those weaknesses, and preparing for grad school.

**SN: How does EDGE encourage community building among participants?**

**TFN:** I personally think that there is something to be said for overcoming difficult things together that builds community, and EDGE successfully balances difficult things with pleasure. That balance really promotes a healthy learning environment.

In the first week of the program, we had something called "Difficult Dialogues." During this time, trained professionals came in and had deep conversations with us about our worries for graduate school and past traumas as women or people of color in the mathematical sciences. Having that dialogue in the very beginning allowed us to see each other as humans and acknowledge the difficulties that we had faced. Exposing each other to those kinds of experiences established a sense of community.

**SN: The 2021 program was entirely in person. What was it like to physically gather with your peers after a year and a half of virtual events?**

**TFN:** Honestly it makes me emotional! Being together in person was amazing. Everybody knows that learning or working online is not the same, and I lost a lot of my math confidence over the past year. It was very difficult to communicate online; it's hard to gauge when to speak, and nobody wants to take more time on Zoom because everyone has Zoom fatigue. Being in person kind of diminished all of that. We were

See **EDGE Fellow** on page 5

siam news

ISSN 1557–9573. Copyright 2021, all rights reserved, by the Society for Industrial and Applied Mathematics, SIAM, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688; (215) 382-9800; siam@siam.org. To be published 10 times in 2021: January/February, March, April, May, June, July/August, September, October, November, and December. The material published herein is not endorsed by SIAM, nor is it intended to reflect SIAM's opinion. The editors reserve the right to select and edit all material submitted for publication.

**Advertisers:** For display advertising rates and information, contact Kristin O'Neill at marketing@siam.org.

**One-year subscription (nonmembers):** Electronic-only subscription is free. $73.00 subscription rate worldwide for print copies. SIAM members and subscribers should allow eight weeks for an address change to be effected. Change of address notice should include old and new addresses with zip codes. Please request address change only if it will last six months or more.

**Editorial Board**
H. Kaper, *Editor-in-Chief, Georgetown University, USA*
K. Burke, *University of California, Davis, USA*
A.S. El-Bakry, *ExxonMobil Production Co., USA*
J.M. Hyman, *Tulane University, USA*
O. Marin, *Argonne National Laboratory, USA*
L.C. McInnes, *Argonne National Laboratory, USA*
S. Minkoff, *University of Texas at Dallas, USA*
N. Nigam, *Simon Fraser University, Canada*
A. Pinar, *Sandia National Laboratories, USA*
R.A. Renaut, *Arizona State University, USA*

**Representatives, SIAM Activity Groups**
**Algebraic Geometry**
B. Mourrain, *Inria, France*
**Analysis of Partial Differential Equations**
G.G. Chen, *University of Oxford, UK*
**Applied Mathematics Education**
P. Seshaiyer, *George Mason University, USA*
**Computational Science and Engineering**
S. Rajamanickam, *Sandia National Laboratories, USA*
**Control and Systems Theory**
F. Leve, *Air Force Office of Scientific Research, USA*
**Discrete Mathematics**
D. Hochbaum, *University of California, Berkeley, USA*
**Dynamical Systems**
K. Burke, *University of California, Davis, USA*
**Financial Mathematics and Engineering**
A. Capponi, *Columbia University, USA*
**Geometric Design**
J. Peters, *University of Florida, USA*
**Geosciences**
T. Mayo, *Emory University, USA*
**Imaging Science**
G. Kutyniok, *Technische Universität Berlin, Germany*
**Life Sciences**
M.A. Horn, *Case Western Reserve University, USA*
**Linear Algebra**
R. Renaut, *Arizona State University, USA*
**Mathematical Aspects of Materials Science**
K. Bhattacharya, *California Institute of Technology, USA*
**Mathematics of Planet Earth**
R. Welter, *Boston University, USA*
**Nonlinear Waves and Coherent Structures**
K. Oliveras, *Seattle University, USA*
**Optimization**
A. Wächter, *Northwestern University, USA*
**Orthogonal Polynomials and Special Functions**
P. Clarkson, *University of Kent, UK*
**Uncertainty Quantification**
E. Spiller, *Marquette University, USA*

***SIAM News* Staff**
L.I. Sorg, *managing editor, sorg@siam.org*
J.M. Kunze, *associate editor, kunze@siam.org*

**Printed in the USA.**
siam *is a registered trademark.*

# AN21 Panel Guides Applied Mathematicians Seeking Careers in Business, Industry, and Government

*By Jillian Kunze*

A wide variety of exciting career paths in business, industry, and government (BIG) are available to applied mathematicians and computational and data scientists. During the 2021 SIAM Annual Meeting,[1] which took place virtually in July, an industry panel discussed strategies for effectively pursuing these opportunities. The panel was chaired by Sharon Arroyo (The Boeing Company), moderated by Kevin Bongiovanni (Raytheon Systems), and organized by the SIAM Industry Committee.[2] Panelists Marylesa Howard (Nevada National Security Site), Cosmin Ionita (MathWorks), Penporn Koanantakool (Google, Inc.), Lois Curfman McInnes (Argonne National Laboratory), Raymond Perkins (Facebook), and Nessy Tania (Pfizer, Inc.) encouraged attendees to consider BIG careers and provided advice for current and future job seekers.

Each panelist related their own unique experiences within BIG settings. For instance, McInnes noted that U.S. Department of Energy national laboratories[3] offer many interesting positions in various domains of science and engineering for researchers to collaborate on impactful and reusable projects. Perkins realized in graduate school that his interests were more aligned with business than academia and has thus worked to leverage his quantitative background in industry. "Try to expose yourself as much as possible to opportunities," he said.

CAREERS IN MATHEMATICAL SCIENCES

Tania, who transitioned from academia to the biopharmaceutical sector, also urged attendees to explore their options. "Really think about what you want to do and imagine that there are different models of success," she said. "Ask yourself early and often about your passions and what you like about your day-to-day job." Ionita echoed this sentiment and added that anyone who has utilized the tools of applied mathematics can find opportunities in BIG. "There are fruitful careers in industry as an applied mathematician, even though your degree might not say 'applied mathematics,'" he said.

The daily tasks of BIG mathematicians often vary widely, both between different individuals and over the course of a single career. Depending on the organization, positions may involve varying levels of collaboration, individual focus, and coordination with groups of people. In some cases, employees can tailor this configuration to their own interests. For example, there is a great deal of variety and flexibility in Howard's average work day, though communicating with other professionals always remains a major component of her job. "I spend a lot of time not talking math, or trying to take math and put it into a language that other people will understand," she said. "This allows me to communicate my ideas effectively in a physics sense or a management sense."

Those who are pursuing BIG careers are sometimes intimidated because many jobs that are appropriate for applicants with applied mathematics backgrounds are not titled as such. Tania advised students to visit the career centers at their universities to learn about jobs for which they might be qualified and find information on the career paths of alumni from their departments. Career centers may even be able to facilitate virtual meetings between current students and alumni. In addition, the BIG Math Network[4] provides articles that explain the role of applied mathematicians in industry, and SIAM's career page lists a number of BIG organizations[5] that hire mathematicians. Career fairs are similarly valuable. During the next SIAM Career

*See* **Careers** *on page 6*

1 https://www.siam.org/conferences/cm/conference/an21
2 https://www.siam.org/about-siam/committees/industry-committee
3 https://www.energy.gov/national-laboratories
4 https://bigmathnetwork.org
5 https://www.siam.org/careers/resources/details/companies-and-industries

## Digital Twins

*Continued from page 1*

which is informed by observational data $O_t$. One can use the computational models that comprise the digital twin to predict quantities of interest $Q_t$, which in turn inform control inputs $U_t$ that influence the asset state. Definition of a reward $R_t$ facilitates the analysis and optimization of the system. Here, the subscript $t$ denotes a discrete time step.

A graphical model serves as a formal mathematical representation of the way in which each of these aforementioned elements interact with one another and how they evolve over time. Such models can take many specific forms, including Bayesian networks, factor graphs, or deep generative networks. Figure 3 illustrates a dynamic decision network—specifically a dynamic Bayesian network with the addition of decision nodes—for a digital twin system. Each quantity is modeled at a particular instant in time as a random variable, which is represented as a node in the graph. Edges signify relationships between quantities; these are typically encoded as conditional probability distributions. For example, the conditional probability

$$p\,(D_t\,|\,D_{t-1}=d_{t-1},\,O_t=o_t)$$

models the transition in the digital state from timestep $t-1$ to timestep $t$, conditioned on observed data $o_t$. Physics-based models that constrain the system dynamics according to physical governing equations may be embedded within the definition of this probability distribution. The distribution could also be data-driven, e.g., fitted to historical asset data.

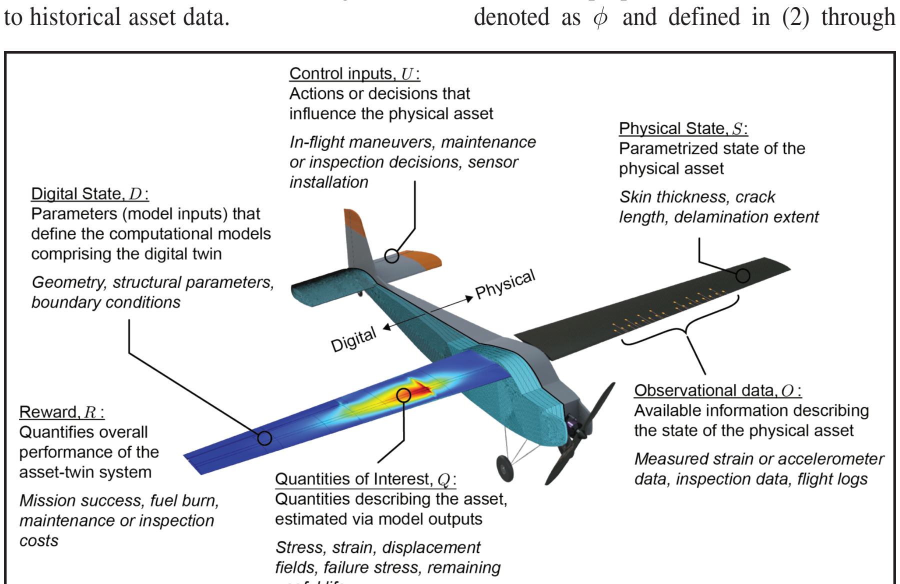



**Figure 2.** *The elements that define a mathematical model of an unmanned aerial vehicle and its associated digital twin. Figure adapted from [3].*

The graph's sparse connectivity encodes a set of known or assumed conditional independencies. One can exploit this conditional independence structure to factorize joint distributions over variables of interest in the model. For example, we can factorize our belief about the digital state $D_t$, quantities of interest $Q_t$, and rewards $R_t$, conditioned on observed variables for all timesteps until the current time (namely the data $O_t=o_t$ and enacted control inputs $U_t=u_t$ for $t\in\{0,\dots,t_c\}$), according to the structure of the proposed graphical model as

$$p(D_0,\dots,D_{t_c},Q_0,\dots,Q_{t_c},R_0,\dots,R_{t_c}\,|\,o_0,\dots,o_{t_c},u_0,\dots,u_{t_c})=\prod_{t=0}^{t_c}[\phi_t^{\text{update}}\phi_t^{\text{QoI}}\phi_t^{\text{evaluation}}], \quad (1)$$

where

$$\phi_t^{\text{update}}=p(D_t\,|\,D_{t-1},U_{t-1}=u_{t-1},O_t=o_t) \quad (2)$$

$$\phi_t^{\text{QoI}}=p(Q_t\,|\,D_t) \quad (3)$$

$$\phi_t^{\text{evaluation}}=p\,(R_t\,|\,D_t,R_t,Q_t,U_t=u_t,O_t=o_t). \quad (4)$$

One can formulate predictions in a similar manner by extending this belief state to include digital state, quantity of interest, and reward variables up until the chosen prediction horizon $t_p$.

Factored representations such as (1) serve two purposes. First, each factor—denoted as $\phi$ and defined in (2) through (4)—is a conditional probability distribution that exposes particular processes or interactions that must be characterized in a digital twin (e.g., asset state dynamics, data generation and observation, computational estimation of quantities of interest, and so forth). Second, the factorization serves as a foundation for deriving efficient sequential Bayesian inference algorithms that leverage the digital twin models to enable high-level digital twin functionality like asset monitoring, prediction, and optimization [3].

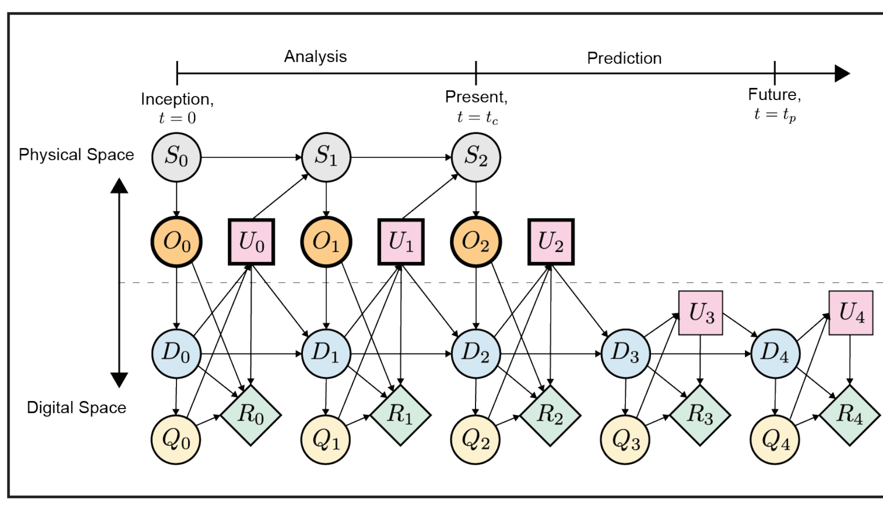



**Figure 3.** *A probabilistic graphical model that describes the coupled evolution of a physical asset and its associated digital twin. Figure adapted from [3].*

### Outlook

The early successes of digital twin deployment point to the idea's value and potential impact. Now is the time for the applied mathematics and computational science communities to develop the rigorous mathematical underpinnings and scalable algorithms that will take digital twins to the next level. As noted in [4], for inspiration we can look to the evolution of the finite element method from an expert-driven approach that required specialization for each different application to a broadly applicable analysis and design tool that is now in the hands of every engineer. This evolution has been enabled by foundational mathematical theory, computing scalability achieved through a combination of hardware and algorithmic advances, and flexible software implementations.

In order to advance digital twins to a similar level of maturity and accessibility, our community's work in a variety of areas—including physics-based modeling, inverse problems, data assimilation, uncertainty quantification, optimal control, optimal experimental design, surrogate modeling, scientific machine learning, scalable algorithms, and scientific software—has an important role to play.

*This article is based on Karen Willcox's invited talk at the 2021 SIAM Conference on Computational Science and Engineering (CSE21),[1] which took place virtually in March.*

### References


[1] AIAA Digital Engineering Integration Committee. (2020). *Digital twin: Definition & value*. (AIAA and AIA Position Paper). American Institute of Aeronautics and Astronautics and Aerospace Industries Association.

[2] Dallaert, F. (2021). Factor graphs: Exploiting structure in robotics. *Ann. Rev. Cont. Robot. Auton. Syst.*, *4*(1), 141-166.

[3] Kapteyn, M., Pretorius, J., & Willcox, K. (2021). A probabilistic graphical model foundation for enabling predictive digital twins at scale. *Nat. Comput. Sci.*, *1*(5), 337-347.

[4] Niederer, S., Sacks, M., Girolami, M., & Willcox, K., (2021). Scaling digital twins from the artisanal to the industrial. *Nat. Comput. Sci.*, *1*(5), 313-320.


*Michael G. Kapteyn is a postdoctoral fellow in the Oden Institute for Computational Engineering and Sciences at the University of Texas at Austin. He recently received his Ph.D. in computational science and engineering from the Massachusetts Institute of Technology, where he worked in the Aerospace Computational Design Laboratory developing mathematical and computational foundations for digital twins. Karen E. Willcox is director of the Oden Institute for Computational Engineering and Sciences, Associate Vice President for Research, and professor of aerospace engineering and engineering mechanics at the University of Texas at Austin. She is also an external professor at the Santa Fe Institute.*

1 https://www.siam.org/conferences/cm/conference/cse21

## Bird Lungs

*Continued from page 1*

us to assess the way in which network geometry affects flow rectification. Specifically, our study aims to isolate the effects of the two loops that are connected by junctions. It initially seemed impossible for the loopy network to generate a DC flow. After all, if the network represents the flow of electrical current in a circuit that is modeled by a linear system of ordinary differential equations, one could readily conclude that no DC component exists. But flows in bird lungs are sufficiently fast to involve nonlinear fluid dynamical effects whose consequences in terms of flow patterns are still unknown.

Experiments that were conducted in the Applied Mathematics Laboratory at New York University's Courant Institute of Mathematical Sciences involved a simplified model lung that was made of tubing filled with water. While the water near the piston simply sloshed back and forth, a DC flow spontaneously emerged in the upper loop — as long as the driving amplitude was sufficiently large. This directional flow represents the movement of fresh air in bird lungs that continuously ventilates oxygen-absorbing tissues.

While these experiments demonstrated that our simplified model generates directed flows, the fluid dynamical mechanism remained unclear. To dig deeper, we conducted numerical simulations of the two-dimensional incompressible Navier-Stokes equations with no-slip boundary conditions on the piston and the walls of the domain (see Figure 1c, on page 1). The equations are

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t}+\boldsymbol{u}\cdot\nabla\boldsymbol{u}\right)=-\nabla p+\mu\Delta\boldsymbol{u},$$

$$\nabla\cdot\boldsymbol{u}=0.$$

Here, $\boldsymbol{u}(\boldsymbol{x},t)$ and $p(\boldsymbol{x},t)$ are the fluid velocity and pressure respectively, and $\rho$ and $\mu$ are the fluid density and dynamic viscosity. The left side of the first equation represents the fluid inertia and the two terms on the right side respectively account for pressure and viscous forces. We conducted the simulations using the finite element method and employed an Arbitrary Lagrangian-Eulerian formulation to treat the neighborhood of the moving piston.

These simulations allowed us to visualize the flow fields, which revealed a subtle interplay between inertial and viscous effects at the T-shaped junctions. Figure 1d (on page 1) highlights the key events. At peak velocity during inhalation (the first image in Figure 1d), fluid is injected from the right via the blue arrow and predominantly moves straight past the junction, with minimal turning up the side branch. This movement matches the intuition that the flow's inertia causes it to maintain a straight course. Viscosity compels the flow to separate from the corner, resulting in a vortex that is shed and then "plugs" the side branch. At peak velocity during exhalation (the second image in Figure 1d), fluid exits the junction at the red arrow and draws more equally from the two branches because the vorticity that is produced at the corner "hugs" the wall instead of detaching to form a plug. Integrating in time over the course of a cycle, the indicated junction thus has a net flux that corresponds to circulation in the clockwise sense around the network's upper loop.

What permits this rectification/pumping to occur without conventional valves? One key ingredient is that the fluid and operating conditions must allow both inertial and viscous effects to participate. Rectified flows of the type that we see here are impossible in viscous-dominated conditions where inertia is neglected; the governing Stokes equation is reversible and the so-called scallop theorem [8] ensures that flows induced in one stroke are retraced reversely in the return stroke. Kelvin's circulation theorem means that rectification is also precluded in the other extreme of inviscid and irrotational flow (potential flow) [3]. In the parlance of fluid dynamics, such valveless rectification fails at the extremes of the Reynolds number, $\mathrm{Re}=0$ and $\mathrm{Re}=\infty$. Fortunately, there is a lot of room in between where the mechanism is operable!

Just as important are the junctions' geometrical anisotropy and their connectivity to form the loops around which circulation develops. The junctions have distinct side and straight branches that tend to bias incoming flows. We do not believe that the T-shape in our study is critical, but some degree of anisotropy is important. The junctions must also be "wired" in a way that exploits the tendency of inertial flows to go straight. In our model, the straight branch of each junction feeds into the side branch of the other to form a loop that is then fed by both straight segments. One-way flows occur as long as we maintain this property, which we confirmed in additional experiments on networks with various geometries.

Our study focuses specifically on network topology and the nonlinear flows at junctions, but researchers have much more to do and understand. While there is beauty and power in the investigation of idealized systems, one cannot help but wonder about the roles of the many other lung complexities. This line of reasoning inspires mathematical questions about the optimization of junction shape, network structure, and kinematic forcing to enhance directed flows. Like all good mysteries, the unexpectedly fascinating aerodynamics of bird respiration endures. Our study has uncovered some clues but raises as many questions as answers.

### References


[1] Banzett, R.B., Nations, C.S., Wang, N., Fredberg, J.J., & Butler, J.P. (1991). Pressure profiles show features essential to aerodynamic valving in geese. *Respir. Physiol.*, *84*(3), 295-309.

[2] Bretz, W.L., & Schmidt-Nielsen, K. (1971). Bird respiration: Flow patterns in the duck lung. *J. Exp. Biol.*, *54*(1), 103-118.

[3] Childress, S. (2009). *An introduction to theoretical fluid mechanics*. New York, NY/Providence, RI: Courant Institute of Mathematical Sciences and the American Mathematical Society.

[4] Duncker, H.-R. (1974). Structure of the avian respiratory tract. *Respir. Physiol.*, *22*(1-2), 1-19.

[5] Maina, J.N. (2000). What it takes to fly: The structural and functional respiratory refinements in birds and bats. *J. Exp. Biol.*, *203*(20), 3045-3064.

[6] Maina, J.N. (2017). Pivotal debates and controversies on the structure and function of the avian respiratory system: Setting the record straight. *Biol. Rev. Camb. Philos. Soc.*, *92*(3), 1475-1504.

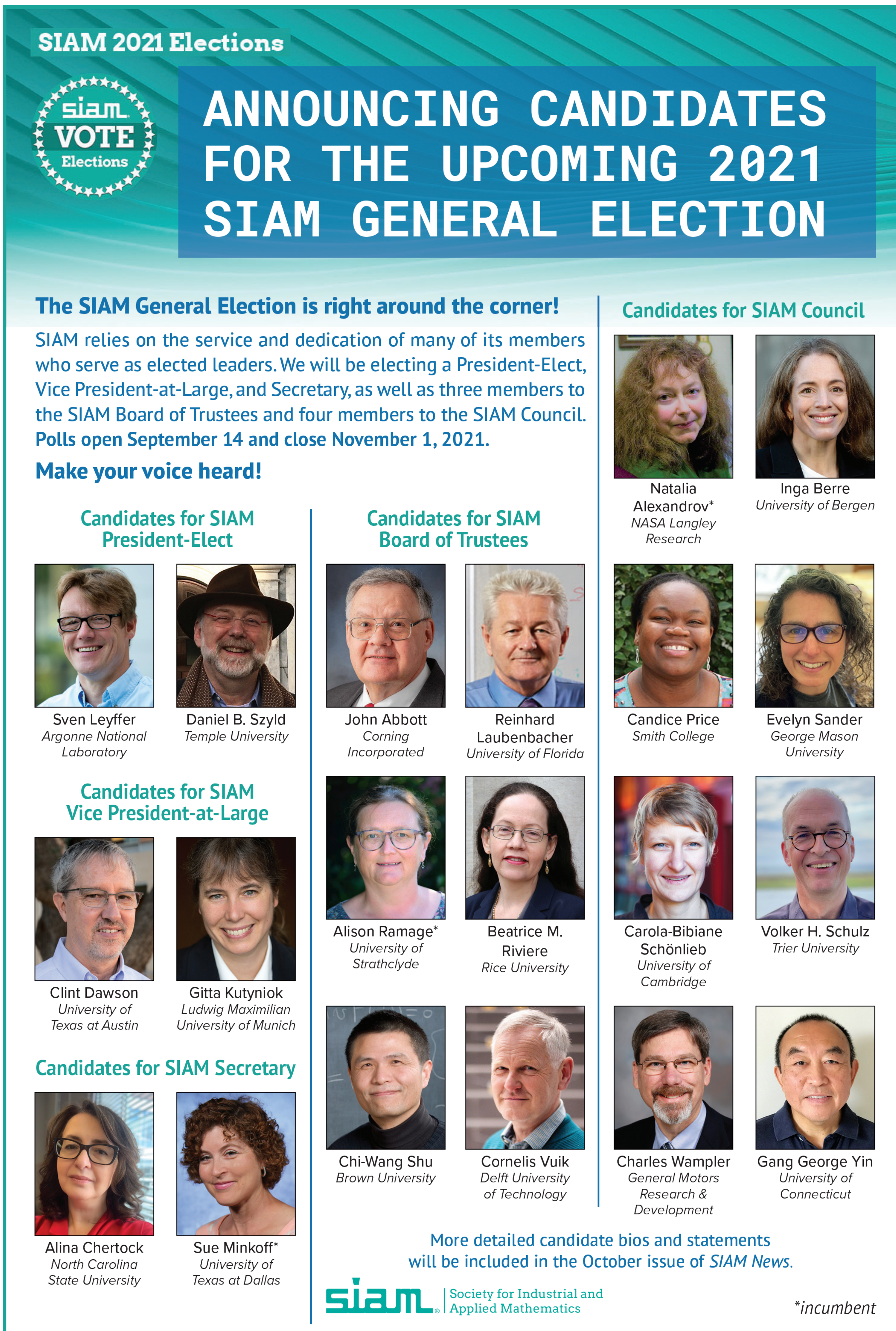
[7] Nguyen, Q.M., Oza, A.U., Abouezzi, J., Sun, G., Childress, S., Frederick, C., & Ristroph, L. (2021). Flow rectification in loopy network models of bird lungs. *Phys. Rev. Lett.*, *126*(11), 114501.

[8] Purcell, E.M. (1977). Life at low Reynolds number. *Am. J. Phys.*, *45*(1), 3-11.

[9] Scott, G.R. (2011). Elevated performance: The unique physiology of birds that fly at high altitudes. *J. Exp. Biol.*, *214*(15), 2455-2462.

[10] Tucker, V.A. (1968). Respiratory physiology of house sparrows in relation to high-altitude flight. *J. Exp. Biol.*, *48*(1), 55-66.


*Anand U. Oza is an applied mathematician at the New Jersey Institute of Technology (NJIT) who works on fluid mechanics and nonlinear dynamics, with applications to soft matter physics and biological systems. Christina Frederick is an applied mathematician at NJIT with a focus on mathematical analysis and computational modeling in physics and engineering. Leif Ristroph is an applied mathematician and experimental physicist who directs the Applied Mathematics Laboratory at New York University's Courant Institute of Mathematical Sciences.*

# Prime Gap Breakthrough

**Bounded Gaps Between Primes: The Epic Breakthroughs of the Early Twenty-First Century**. *By Kevin Broughan. Cambridge University Press, Cambridge, U.K., April 2021. 590 pages, $49.99.*

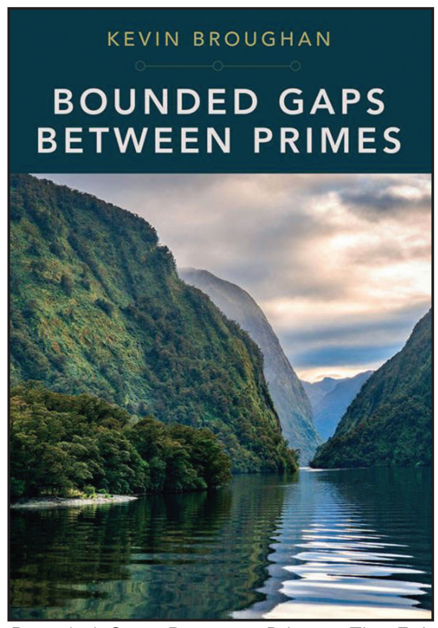
BOOK REVIEW
*By James Case*

In *Bounded Gaps Between Primes*, Kevin Broughan of the University of Waikato has written a remarkably comprehensive account of the 2013 landmark discovery that demonstrated the existence of infinitely many disjoint real intervals of length at most $L$ that contain two or more prime numbers. The originally reported length was $L=7\times10^7$, but follow-up efforts reduced that figure to 246 by 2014. Some have noted that proving the Elliott-Halberstam conjecture would cause a further reduction to $L=6$.

Yet both $L=246$ and $L=6$ are still far from $L=2$, which would settle the twin primes conjecture that Alphonse de Polignac first enunciated in 1849. If nothing else, the new results appear to suggest that the long-unapproachable conjecture may be settled in the not-too-distant future.

Broughan provides a table of 17 publications on prime gaps, the first 11 of which appeared between 1923 and 1988. From 2006 to 2009, two more papers introduced new methods but failed to eliminate the possibility that $\liminf_n (p_{n+1}-p_n)=\infty$, where $p_n$ is the $n^{\text{th}}$ prime in the list that begins with $p_1=2$. The last four tabulated papers—published in 2013 and 2015—reduced $L$ successively from $7\times10^7$ to 4,680, then to 600, and finally to 246. The authors were Yitang Zhang, a collective known as Polymath8a, James Maynard and Terence Tao, and a second collective known as Polymath8b.

In 2009, Timothy Gowers proposed the Polymath idea in a blog entry entitled, "Is massively collaborative mathematics possible?"[1] He suggested that anyone with thoughts about a specific problem should speak up, even if their ideas are undeveloped or likely wrong. After all, the knowledge that a large group can bring to a given problem is considerably greater than the knowledge of just one or two individuals. Gowers even proposed a short list of problems for which massive collaboration seemed suitable.



*Bounded Gaps Between Primes: The Epic Breakthroughs of the Early Twenty-First Century. By Kevin Broughan. Courtesy of Cambridge University Press.*

Although Broughan traces the twin primes conjecture back to its apparent roots, he focuses on developments from the present century. *Bounded Gaps Between Primes* offers brief biographical sketches of the main contributors to prime gaps literature and directs interested readers to more extensive interviews that are available elsewhere.

[1] https://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible

By far the most poignant sketch is that of Zhang, who obtained the original estimate $L\le7\times10^7$. Born in 1955, Zhang grew up in mainland China during the Cultural Revolution. Unable to go to school, he taught himself with whatever books he could find and later attended Peking University, where he became a star student. Zhang came to the U.S. in the 1980s and earned a Ph.D. from Purdue University in 1991. But the job market was unwelcoming and he worked odd jobs to support himself until 1999, when he became an adjunct professor at the University of New Hampshire teaching introductory calculus. Through it all, Zhang's mathematical curiosity remained unrelenting.

Hardy and Littlewood's 1923 paper is the first entry in Broughan's table; this publication proved the existence of infinitely many values of $n\le x$ for which

$$p_{n+1}-p_n \le \left(\frac{2}{3}+o(1)\right)\log x. \qquad (1)$$

The constant $2/3$ was subsequently improved numerous times before Daniel Alan Goldston, Yoichi Motohashi, János Pintz, and Cem Yalçın Yıldırım (GMPY) replaced the right side of (1) first with $o(\log x)$ and then with $C(\log x)^{\frac{1}{2}}(\log\log x)^2$ for an appropriate constant $C$. Hardy and Littlewood also conjectured that

$$\pi_2(x)\sim C'x/(\log x)^2 \qquad (2)$$

for an appropriate $C'$, where $\pi_2(x)$ is the number of primes $p\le x$ such that $p+2$ is also a prime. They were led to this conjecture by Cramér's suggestion that a randomly chosen natural number between 1 and $x$ should be prime with probability $1/\log x$. That result, together with an observation by L.E. Dickson, yielded a significant generalization.

Dickson noticed that certain patterns among the primes occasionally repeat themselves and conjectured that some do so infinitely often. The patterns that he noticed take the form $\mathcal{H}=\{h_1,h_2,\dots,h_k\}$ and are such that numerous translates $\mathcal{H}+n=\{h_1+n,h_2+\ n,\dots,h_k+n\}$ consist exclusively of primes. $\mathcal{H}$ must satisfy certain conditions to behave in this manner. For instance, all of the differences $h_i-h_j$ must be even because if even one difference were odd, $\mathcal{H}$ and every translate of $\mathcal{H}$ would contain an even (nonprime) number. Dickson separated the set of all possible patterns $\mathcal{H}$ into two subsets: an "inadmissible" subset that could never perform in the proposed manner and an "admissible" subset that conceivably might. The pattern $\mathcal{H}=\{0,2\}$ belongs to Dickson's admissible set and corresponds to the twin primes conjecture. Zhang's proof that $L\le7\times10^7$ utilizes an $\mathcal{H}$ of dimension $k=3.5\times10^6$.

The so-called Dickson-Hardy-Littlewood conjecture generalizes the original Hardy-Littlewood conjecture by asserting that

*See* **Prime Gap** *on page 7*

## EDGE Fellow

*Continued from page 2*

able to think out loud for the first time in over a year, and it really helped to accelerate a lot of my thought processes.

**SN: SIAM President Susanne C. Brenner visited EDGE students on site this year. Did you have an opportunity to speak with her about SIAM?**

**TFN:** We had the opportunity to speak with Dr. Brenner during one of our colloquium sessions. She emphasized the many opportunities for students in the SIAM community and encouraged us to have a go-getter mentality at conferences. Through this conversation, I gained some perspective on how mathematicians might navigate conferences like the SIAM Annual Meeting. My SIAM membership will allow me to participate in workshops and events with other mathematicians in my fields of interest and explore the many research areas I am currently enamored by.

If a graduate student in the mathematical sciences wants to be at the forefront of a particular field, joining SIAM would be a great first step. Membership offers them the potential to tap into this ecosystem of collaboration, career advancement, educational resources, and more. The value also goes both ways — the advancement of any field heavily depends on the inclusion of all who are dedicated to the mission.

**SN: What was your initial reaction upon learning that you are the inaugural SIAM EDGE Fellow?**

**TFN:** I was so surprised and really flattered. I feel sort of like a spokesperson for EDGE because I truly believe in its mission statement. It gave me the recharge that I needed to feel confident in my math ability, but also confident in myself as a student and as someone who can add value to the mathematical sciences. I really hope that I can be a mentor someday or even a professor teaching the classes. That would be a dream, and I will always be an EDGE cheerleader.

**SN: You just began graduate classes in applied mathematics at Cornell University. Do you have any expectations for graduate school?**

**TFN:** This is still a hard question for me to answer because my interests are so vast right now! I'm currently letting the classes I enjoy the most guide my way. It's going to be a lot of experimenting with and gauging the types of courses I take, the people I talk to, and the relationships I form, and letting these things influence the kinds of projects that I pursue. I would like to conduct research that involves optimization—maybe combinatorial optimization in particular—and I hope to develop optimization techniques or utilize new methods in the field to build, restore, and/or nurture sustainable cities.

**SN: What are your career aspirations for life after graduate school?**

**TFN:** I want to stay in academia and work for a teaching university. Teaching is where my heart is, and solving really hard questions with friends is the goal. With EDGE, I had a month of making friends and doing math. It was stressful but it's what I want to go to grad school for: to solve math problems with friends. It was exactly what I needed.

**SN: What advice do you have for undergraduate students who are thinking about applying to EDGE?**

**TFN:** Anyone who is a woman in mathematics should take the opportunity to meet other inspiring women and get paid to do so. That is the best thing you can do for yourself — to build not just your network, but your allies in a community where there are not many people like you. Math can be really isolating because you do a lot of thinking on your own, but EDGE offers a chance for you to have friends at different universities around both the country and world. It's an opportunity to know a face when you go to conferences, and that in itself is a great networking tool. There are a lot of benefits, but there is also a lot that you can offer by being there and adding your perspective. Everybody in the individual cohort adds some sort of viewpoint; your perspective is needed and should be heard.

*Lina Sorg is the managing editor of SIAM News.*

## Careers

*Continued from page 3*

Fair,[6] which will take place virtually on November 9, employers will recruit for job and internship opportunities.

Conferences, workshops, and internships like Research Experiences for Undergraduates[7] are also excellent ways to explore BIG careers and find job offerings. Poster sessions at in-person conferences are often especially beneficial for networking purposes. "Sometimes I ask students directly if they would be interested in an internship after chatting about their posters," Ionita said. Furthermore, speaking to recruiters if the chance arises—no matter one's level of interest in a job—can effectively build a network.

Another reliable method for identifying opportunities in BIG is to peruse the listings at professional societies, such as the SIAM Job Board.[8] "Start paying attention to those advertisements," Tania said. "It might be that you're not on the job market yet, but start looking at the requirements and what types of jobs people are looking for." Monitoring current job postings can help students who are seeking BIG careers focus on transferable skills and tailor their resumes for positions of interest.

When applying for openings in BIG, one should highlight a strong understanding of probability and statistics, a background in linear algebra, and some form of coding knowledge. "When I hire someone, I don't care what coding languages they know," Howard said. "Just the fact that they understand the logical process of programming [is enough]." The ability to communicate across disciplines and clearly express ideas both verbally and in writing is also essential, as is a willingness to continue to learn. "Open your mind to people who have different training and backgrounds than you as an applied mathematician, and who describe things using different terminology that you invest time in to understand and learn," McInnes said. Participation in a SIAM Student Chapter[9] can showcase one's interpersonal skills, including leadership capability and teamwork experience.

Many BIG job employers expect applicants to demonstrate their software skills, and Perkins recommended creating a personal website or GitHub to show off past projects. Making software projects open source if possible allows prospective employers to view the code, confirms that people actually use the work, and counts the number of users who downloaded it. "Building a community that uses it, posts comments, and helps maintain it will show that you can develop something that people really need," Koanantakool said. If making open-source code is not feasible, one might consider getting involved in a user community for existing software and providing input to the development team.

[6] https://go.siam.org/careerfair2021
[7] https://www.nsf.gov/crssprgm/reu
[8] https://www.siam.org/careers/job-board
[9] https://www.siam.org/students-education/student-chapters

Applicants can still exhibit relevant skills even if they lack direct work experience in fields such as data science or machine learning. Hackathons and online competitions—like the machine learning competitions that are hosted on kaggle[10]—serve as examples of clear experience on a resume. "You need to somehow demonstrate to the recruiters and folks who are reading your resume that you actually know how to do a data scientist's job," Koanantakool said. A certification from an online course like Coursera[11] or a boot camp program also bolsters an application.

During the interview process, prospective employees must be ready to draw connections between their backgrounds and the opening's application area that are not immediately apparent. "Sometimes this requires you to think creatively," Perkins said. "You really have to consider what that network of mathematics looks like. If you get the interview and have the skillset and understanding, you'll be able to demonstrate that." Expressing genuine interest in the project in question and speaking about it intelligently can convince hiring managers of one's suitability.

[10] https://www.kaggle.com
[11] https://www.coursera.org

*A panel at the 2021 SIAM Annual Meeting, which took place virtually in July, outlined the preparation and application process for applied mathematicians who are seeking employment in business, industry, and government (BIG). Top row, left to right: Marylesa Howard (Nevada National Security Site), Cosmin Ionita (MathWorks), and Penporn Koanantakool (Google, Inc.). Bottom row, left to right: Lois Curfman McInnes (Argonne National Laboratory), Raymond Perkins (Facebook), and Nessy Tania (Pfizer, Inc.).*

To close out the session, panelists reminded students to actively pursue internships and similar opportunities that will help them explore BIG and ultimately decide on a career path. "Applied mathematics is such an amazing foundation for a career that's exciting and enriching," McInnes said. "It can go so many directions depending upon your interests, so focus on what you as an individual are passionate about. Don't feel that you need to follow the model of anybody else; really be true to yourself and seek out what you enjoy."

*Jillian Kunze is the associate editor of SIAM News.*


### Want to Place a Professional Opportunity Ad or Announcement in *SIAM News*?

Please send copy for classified advertisements and announcements in *SIAM News* to **marketing@siam.org**.

For details, visit **www.siam.org/advertising**.

Check out the SIAM Job Board at **jobs.siam.org** to view all recent job postings.

## Prime Gap

*Continued from page 5*

$$\pi(x,\mathcal{H}) = \#\{n \le x : (n+h_1,\ldots,n+h_k) \in \mathbb{P}^k\} \sim g(\mathcal{H}) \cdot x/(\log x)^k, \quad (3)$$

where $\mathbb{P}$ denotes the collection of all primes and $g(\mathcal{H})$ is readily computable. If confirmed, this would imply both the twin primes conjecture and the fact that $p+\{0,2,6\}$ and $p+\{0,4,6\}$ consist of primes alone for infinitely many primes $p$. However, $p+\{0,2,4\}$ is known to consist exclusively of primes for at most a finite number of $p \in \mathbb{P}$. And for every $k \in \mathbb{N}$, there are infinitely many $k$-tuples $\mathcal{H}$ in Dickson's admissible set.

From time immemorial, *sieves* have been the tool of choice for the study of primes. Broughan suspects that a general theory of sieves—with applications to branches of mathematics beyond number theory—might someday exist. At the moment, however, there are many different sieves. In addition to the sieve of Eratosthenes and Adrien-Marie Legendre's variation, *Bounded Gaps Between Primes* mentions eight important sieves — most of which have useful variants.

Modern sieves estimate $S(A,\mathcal{P},z) = \#\{a \in A : (a,\mathcal{P}(z)) = 1\}$. Here, $(a,b)$ denotes the greatest common divisor of $a$ and $b$, $\mathcal{P}$ is an infinite set of primes, $\mathcal{P}(z)$ is the product of all primes in $\mathcal{P}$ that do not exceed $z$, and $A = \{a_1,\ldots,a_n\}$ is a finite sequence of natural numbers. If one "sifts" $A$ by removing all elements that share a prime factor with $\mathcal{P}(z)$, $S(A,\mathcal{P},z)$ is the number of surviving elements of $A$. If $A_q = \{a_n \in A : q \mid a_n\}$ while $\mu(\cdot)$ is the so-called Möbius function, one may write

$$S(A,\mathcal{P},z) = \sum_{q \mid \mathcal{P}(z)} \mu(q)\,|A_q|. \quad (4)$$

This is an exact version of the Brun sieve. In the special case $\mathcal{P} = \mathbb{P}$—the set of all prime numbers—set $S(A,z) := S(A,\mathbb{P},z)$. Then if $A = \mathbb{N} \cap [1,x]$ and $\sqrt{x} \notin \mathbb{P}$,

$$S(A,\sqrt{x}) = \pi(x) - \pi(\sqrt{x}). \quad (5)$$

This is Legendre's form of the sieve of Eratosthenes, obtained by appealing to the combinatorial principle of inclusion and exclusion.

Mathematician Viggo Brun used his sieve to attack both the twin primes conjecture and the older Goldbach conjecture. By the end of the decade, he was able to show that the sum of the reciprocals of the twin primes converges and that there exist infinitely many natural numbers $n$, such that both $n$ and $n+2$ have at most nine prime factors. While Brun's sieve continues to find occasional applications, Atle Selberg's proposed sieve around 1940 has contributed far more to prime gap breakthroughs.

The distinguishing feature of Selberg's sieve for the estimation of $S(A,\mathcal{P},z)$ is a sequence $\{\lambda_q\}$ that is defined as follows: $\lambda_1 = 1$ and $\lambda_q = 0$ if either $q \ge z$ or $q$ fails to divide $\mathcal{P}(z)$. Otherwise, $\lambda_q$ may be any real number. Since $\mathcal{P}(z)$ is a product of distinct primes, $q < z$ only fails to divide $\mathcal{P}(z)$ if $q$ contains a repeated prime factor. Unless $q$ is "square free," $\lambda_q$ may be any real number. The quantities $\lambda_q$ attracted Selberg's attention because

$$S(A,\mathcal{P},z) \le \sum_{a \in A} \left( \sum_{q \mid (a,\mathcal{P}(z))} \lambda_q \right)^2 = |A|Q + R, \quad (6)$$

where $Q$ is a quadratic form in the $\lambda_q$s. As many as 40 percent thereof are potentially nonzero quantities, which one may choose to minimize the upper bound (6).

GMPY's 2006 paper was celebrated in number theory circles for its use of admissible $k$-tuples, appeal to the Elliott-Halberstam conjecture, and optimization step, all of which have since found applications elsewhere. Zhang's reading of this paper, along with its 2009 sequel, appears to have pointed the way to $L \le 7 \times 10^7$.

Broughan explains that Zhang's result was accepted almost immediately by experts in the field precisely because it did not make extensive use of novel or unfamiliar techniques. Instead, it leveraged a surprising amount of the machinery that mathematicians have developed for the study of prime numbers. For example, the Elliott-Halberstam conjecture involves a parameter $\theta \in [0,1]$; the conjecture is known to be true for $0 \le \theta \le 1/2$ and false for $\theta = 1$. In order to complete his proof, Zhang had to prove it true for $0 \le \theta \le 1/2 + 1/1168$. Figure 1 illustrates the complex nature of Zhang's proof and depicts the logical dependencies among the various theorems and lemmas. Zhang's result is denoted "Thm 5.10." Several of the contributing results are simply variations of those already known to workers in the field. For instance, Theorem 5.8 is Zhang's variation of a theorem due to Bombieri and Vinogradov that seems to underlie every approach to the study of prime gaps.

Everything discussed in this review is contained in the first few chapters of *Bounded Gaps Between Primes*. The later chapters are meant to encompass all content that a novice might wish to know before beginning work in the field. The material is therefore quite technical, precise, and rich in detail. Broughan's treatise will likely remain a standard reference for many years to come.

*James Case writes from Baltimore, Maryland.*

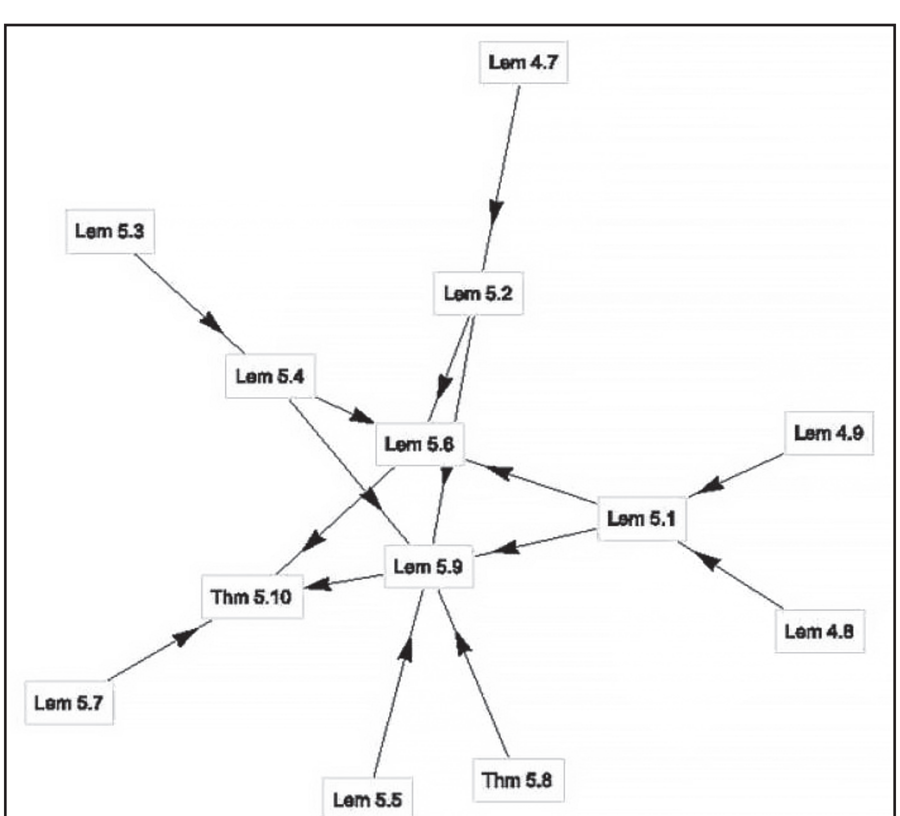



**Figure 1.** *The logical dependencies among the various theorems and lemmas that are involved in Yitang Zhang's proof. Zhang's result is denoted "Thm 5.10." Figure courtesy of Cambridge University Press.*

ICERM

**Institute for Computational and Experimental Research in Mathematics**

**APPLY TO BECOME AN ICERM POSTDOC**

*The Institute for Computational and Experimental Research in Mathematics (ICERM) at Brown University invites applications for its 2022-2023 postdoctoral positions.*

**Postdoctoral Institute Fellows:** ICERM supports two academic-year Postdoctoral Institute Fellows with salary and benefits.

**Postdoctoral Semester Fellows:** ICERM supports five four-month Postdoctoral Fellows each semester with salary and benefits.

**The 2022-2023 Semester Programs are:**
- *Harmonic Analysis and Convexity* (Fall)
- *Modern Discrete Optimization: Mathematics, Algorithms, and Computation* (Spring)

**Eligibility for all ICERM Postdoctoral positions:** Applicants must have completed their Ph.D. within three years of the start of the appointment. Documentation of completion of all requirements for a doctoral degree in mathematics or a related area by the start of the appointment is required.

**For *full consideration:*** applicants must submit an AMS Standard Cover Sheet, curriculum vitae (including publication list), cover letter, research statement, and three letters of recommendation by early January, 2022, to **MathJobs.org** (search under "Brown University").

*Brown University is committed to fostering a diverse and inclusive academic global community; as an EEO/AA employer, Brown considers applicants for employment without regard to, and does not discriminate on the basis of, gender, sex, sexual orientation, gender identity, national origin, age, race, protected veteran status, disability, or any other legally protected status.*

**To learn more** about ICERM programs, organizers, program participants, to submit a proposal, or to submit an application, please visit our website:

**https://icerm.brown.edu**

**Ways to participate:**
*Propose a:*
- semester program
- topical or Hot Topics workshop
- summer undergrad program
- small group research project

*Apply for a:*
- semester program or workshop
- postdoctoral fellowship

*Become an:*
- academic or corporate sponsor

**About ICERM**: The Institute for Computational and Experimental Research in Mathematics is a National Science Foundation Mathematics Institute at Brown University in Providence, Rhode Island. Its mission is to broaden the relationship between mathematics and computation.

121 S. Main Street, 11th Floor
Providence, RI 02903
401-863-5030
info@icerm.brown.edu

NSF

BROWN

# The NIST Big Data Interoperability Framework

*By Wo Chang*

The term "big data" describes the massive amounts of data that are available in the networked, digitized, sensor-laden, and information-driven world. Such data can overwhelm traditional technical approaches, and its continued growth is outpacing scientific and technological developments in data analytics. To help advance ideas in its Big Data Standards Roadmap,[1] the National Institute of Standards and Technology (NIST) created an open community-based Big Data Public Working Group[2] (NBD-PWG). The group aims to develop a secured reference architecture that is both vendor-neutral and technology- and infrastructure-agnostic. The goal is to enable stakeholders (data scientists, researchers, etc.) to perform analytics processing for their given data sources without worrying about the underlying computing environment.

After a multi-year effort, NIST produced nine volumes of special publications under the *NIST Big Data Interoperability Framework* (NBDIF),[3] a collaboration between NIST and more than 90 experts from over 80 industrial, academic, and governmental organizations. Figure 1 displays the relationship between the NBDIF volumes, all of which have specific focuses.

## Big Data Definitions and Taxonomies

Common terms and definitions are critical building blocks for new technologies. This is especially true in the context of big data, which both technical and nontechnical communities tend to describe with "V" terms: volume, velocity, variety, variability, volatility, veracity, visualization, and value. Providing a consensus-based vocabulary allows stakeholders to establish a common understanding of big data. Therefore, NBDIF has defined several terms that are important to the big data industry:

– ***Big data*** consists of extensive datasets—primarily in the characteristics of volume, velocity, variety, and/or variability—that require a scalable architecture for efficient storage, manipulation, and analysis

– ***Big data paradigm*** consists of the distribution of data systems across horizontally coupled, independent resources to achieve the scalability that is needed to efficiently process extensive datasets

– ***Data science*** is the methodology for the synthesis of useful knowledge directly from data through either a process of discovery or hypothesis formulation and hypothesis testing

– ***Data scientist*** is a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and systems engineering to manage the end-to-end data processes in the analytics life cycle.

To fully comprehend the effect of big data, one should examine the granularity of the following components: data elements, related data elements that are grouped into a record that represents a specific entity or event, records that are collected into a dataset, and multiple datasets. Understanding big data characteristics is challenging because the use of parallel big data architectures is based on the interplay of performance, cost, and time constraints on end-to-end system processing. Four fundamental drivers determine the presence of a big data problem:

– ***Volume*** refers to extensive datasets available for analysis to extract valuable information

– ***Velocity*** refers to the measure of the rate of data flow, typically in real-time-like streaming data

– ***Variety*** refers to the need to analyze data from multiple repositories, domains, or types

– ***Variability*** refers to changes in a dataset, whether in the data flow rate, format/structure, and/or volume, thus impacting analytic processing.

## Big Data Use Cases and Requirements

A *use case* is a typical high-level application that extracts requirements or compares usage across fields. For example, NIST created a use case template to develop a consensus list of big data requirements across all stakeholders. To do so, it evaluated the requirements of 51 submitted use cases. The group then extracted 437 specific requirements, aggregated 35 general requirements, and identified seven categories: data source, data transformation, capability, data consumer, security and privacy, life cycle management, and "other." These seven categories helped formulate the big data reference architecture.

[1] https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-7-big-data-standards-roadmap-version-2
[2] https://bigdatawg.nist.gov/home.php
[3] https://bigdatawg.nist.gov/V3_output_docs.php


**Figure 1.** *The Big Data Interoperability Framework (NBDIF) document navigation diagram displays content flow between the nine special publication volumes. Figure courtesy of NIST Big Data Public Working Group.*

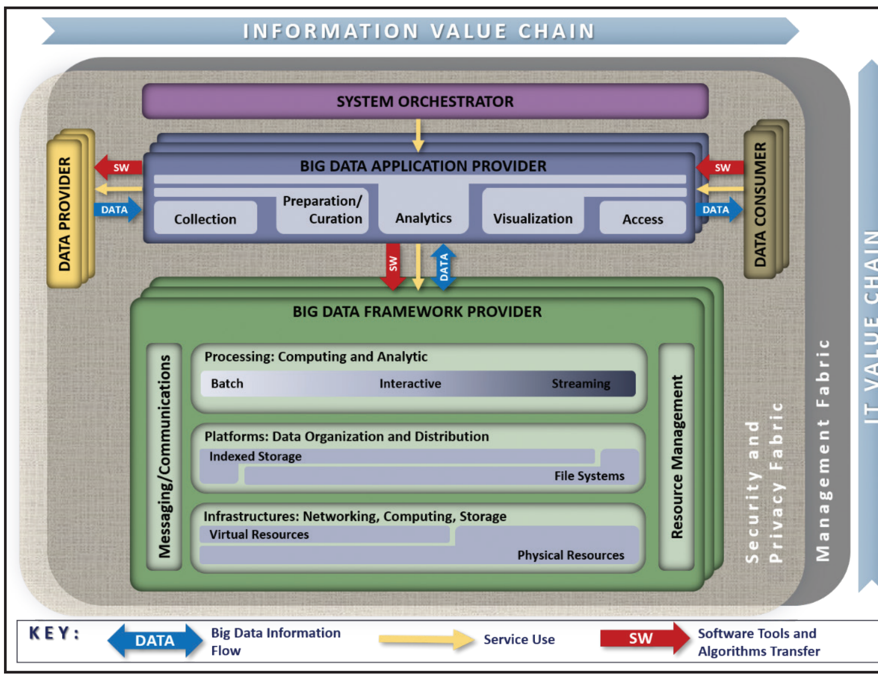



**Figure 2.** *The NIST Big Data Reference Architecture (NBDRA). The blue "DATA" arrows depict the flow of data between NBDRA components, either physically (i.e., by value) or by providing a location and the means for access (i.e., by reference). The red "SW" arrows show the transfer of software tools for big data processing in situ and the yellow "Service Use" arrows represent the application programming interfaces. Figure courtesy of NIST Big Data Public Working Group.*

## Big Data Reference Architecture and Interfaces

The NIST Big Data Reference Architecture (NBDRA)[4] is a high-level conceptual model that aims to facilitate the understanding of big data's operational intricacies. It does not represent the architecture of a specific big data system; instead, NBDRA is a tool for describing, discussing, and developing system-specific architectures via a common framework of reference. The model is not tied to any specific vendor products, services, or reference implementations, nor does it define prescriptive solutions that inhibit innovation.

NBD-PWG identified its key architectural components based on requirements of the derived use cases plus nine industry submissions of big data architectures to develop a reference architecture that is vendor-neutral and technology- and infrastructure-agnostic. NBDRA comprises five logical functional components that are connected by interoperability interfaces. Two fabrics envelop these factors and represent the interwoven nature of both resource management and security and privacy with all five components (see Figure 2). These fabrics provide services and functionality to the five main roles in areas that are specific to any big data solution.

[4] https://www.bigdataframework.org/big-data-architecture

*See* **NIST Big Data** *on page 11*

American Institute of Mathematics

*AIM, the American Institute of Mathematics, sponsors week-long activities in all areas of the mathematical sciences with an emphasis on focused collaborative research.*

**Call for Proposals**

**Workshop Program**

AIM invites proposals for its focused workshop program, both in-person and online. Distinguished by their specific mathematical goals, workshops may involve making progress on a significant unsolved problem or examining the convergence of two distinct areas of mathematics. Workshops are small in size, up to 28 people, to allow for close collaboration among the participants.

**SQuaREs Program**

AIM also invites proposals for the SQuaREs program: Structured Quartet Research Ensembles. More long-term in nature, this program brings together groups of four to six researchers for a week of focused work on a specific research problem in consecutive years.

**Research Communities Program**

AIM is excited to invite proposals for its new Research Communities program. Intended for larger collaborative efforts of 40+ researchers in a virtual setting, these communities receive access to a dedicated online platform with integrated tools to support long-term research collaboration.

More details are available at:

**http://www.aimath.org/research/**

*AIM seeks to promote diversity in the mathematics research community. We encourage proposals which include significant participation of women, underrepresented minorities, junior scientists, and researchers from primarily undergraduate institutions.*

# An Incompetence Problem

MATHEMATICAL CURIOSITIES
*By Mark Levi*

When ruminating recently on the question of finding a person who one can trust to do a job well (speaking of many occupations that range from lawyers and politicians to teachers, plumbers, and physicians), I thought of the following toy problem with no implied connection to reality.

Consider an occupation in which each member is classified as either "competent" or "incompetent" (this idealized binary division excludes gray areas). An incompetent person knows no better and thinks that everyone is competent. On the other hand, a competent individual correctly estimates the proportion $p$ of competents. With the entire population assembled in a conference hall, everybody is asked to drop money in an urn according to their estimates of the percentage of competents in the room. An incompetent thus contributes 1 dollar while a competent contributes $p \leq 1$ dollars, where $p$ is the actual proportion of competents (let us treat $p$ as a continuous variable). Once everyone has chipped in, we find that the average contributed amount is $q$ dollars, where $q \in (0,1]$ (note that $q=0$ is impossible as a consequence of the rules).

The question is as follows: *Can we recover the actual proportion $p$ of competents from the knowledge of $q$?* In other words, *can we filter out the polluting contribution of the incompetents*?

Let us first find the competents' total contribution. With $N$ people in the room, $Np$ of them are competent, each contributing $p$ for the total contribution of $(Np)p = Np^2$. And there are $N(1-p)$ incompetents, each contributing 1 dollar for the total contribution of $N(1-p)$. To summarize,

$$\underbrace{\text{total collected}}_{Nq} = \underbrace{\text{by competents}}_{Np^2} + \underbrace{\text{by incompetents}}_{N(1-p)},$$

or

$$q = p^2 - p + 1 = \left(p - \frac{1}{2}\right)^2 + \frac{3}{4}. \quad (1)$$

From (1), we recover two possible values:

$$p = \frac{1}{2} \pm \sqrt{q - \frac{3}{4}};$$

these are equidistant from the 50 percent mark (see Figure 1). We can think of $q$ as an estimated probability of competence as measured by the imperfect process tainted by the "votes" of the incompetents. Here are a few observations.

**1.** $q \geq 3/4$ always, regardless of the split.

**2.** The minimal $q = 3/4$ occurs when the split is even $(p = 1/2)$. One half (the competents) contributes $1/2$ dollar each and members of the incompetent half each chip in 1 dollar; the average contribution is thus $q = 3/4$ dollars. Figure 1 reflects this conclusion.

**3.** $p$ is determined uniquely only when $q$ is least possible: $q = 3/4$.

**4.** The good-sounding $q = 1$ is not necessarily good; it could also occur due to $p = 0$, where everyone is incompetent and thinks that they and all of their colleagues are great.

**5.** $q > p$ for all values of $p < 1$. This is apparent from (1) or the fact that incompetents contribute the maximum possible amount, thus raising $q$ above $p$.

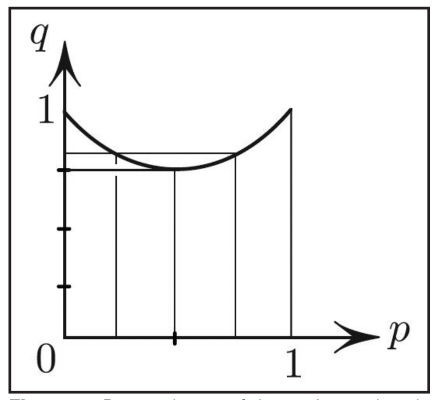



**Figure 1.** *Dependence of the estimated probability ($q$) on the actual probability of competence ($p$). Figure courtesy of Mark Levi.*

After carrying out this thought experiment with no input from reality, I became curious and went online. A cursory search on lawyers yielded estimates of $p = 0.5$ from 1981 [1] and $p = 0.3$ from 2015 [2]. Of course, there is so much ambiguity in these estimates that making conclusions about trends—which do not look good at face value—would be a sign of incompetence itself. Yet regardless of the finer points, the estimate $p \leq 1/2$ is certainly not comforting.

### References


[1] Cramton, R.C. (1981). Lawyer competence and the law schools. *Univer. Arkansas at Little Rock Law Rev.*, *4*(1). Retrieved from https://lawrepository.ualr.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1418&context=lawreview.

[2] Tung, J.R. (2015). There's a 70 percent chance you're incompetent, survey suggests. *FindLaw*. Retrieved from https://blogs.findlaw.com/strategist/2015/10/theres-a-70-percent-chance-youre-incompetent-in-your-law-practice-survey-suggests.html.



*Mark Levi (levi@math.psu.edu) is a professor of mathematics at the Pennsylvania State University.*


# The Why and How of Manuscript Review: What Goes Around Comes Around

*By Reinhard Laubenbacher*

Aside from conducting research itself, communicating it to the broader mathematical sciences community—as well as the general public—is the most important task that a researcher must accomplish, for the sake of progress within both the mathematical sciences and one's own career. It is also the task for which researchers are usually the least prepared. Among other activities, communication often involves publishing results in journals, speaking about them at conferences and in university departments, and describing findings in grant applications that allocate the funds for further research. These mechanisms function like a barter economy. Scholars are willing to travel to other universities and give talks for small honorariums and a dinner because they know that their own departments will host a steady stream of speakers at the same minimal rates. Likewise, they are willing to spend the better part of a week reviewing National Science Foundation (NSF) proposals for a symbolic fee with the understanding that someone else does the same with their own proposals. This exchange prevents the NSF from having to pay reviewers an appropriate consulting fee and allows it to use the funds for additional grants.

In my experience, researchers have many misperceptions when it comes to journals, their different business models, and the responsibility of the research community in their continued operation. Here I intend to clarify the role of journals, the way in which they function, and the actions that individual members of the mathematical sciences research community can take to help journals best serve audience need.

The vast majority of journals are published either by professional societies, such as SIAM or the American Mathematical Society, or commercial publishers like Elsevier or Springer Nature. It is important to understand that both types of publishers aim to make a profit from journals. In fact, professional societies typically depend heavily on publishing revenue for their general operations.

Journals normally have an editor-in-chief who is supported by an editorial board that handles submissions. Editorial board members are rarely compensated for their time and effort, which can be substantial; the same is true of the reviewers that provide their expert opinions on manuscript submissions. Depending on the journal, editorial staff are sometimes available to help process manuscripts until they are handed off to the production department. Journals may receive anywhere from a few hundred to more than a thousand submissions each year.

When an author submits a manuscript, the editor-in-chief determines whether it fits within the scope of the journal and decides if its quality merits further consideration. Next it is assigned to a member of the editorial board who evaluates it further and—if appropriate—invites reviewers. Once all reviews are in, the editorial board member offers a decision or recommendation to the editor-in-chief, who then makes the determination. In some cases, as many as three or four reviews may influence the decision. Of course, there are many possible deviations from this standard procedure. For instance, the editor may seek another reviewer's opinion if initial reviews are divided. If the editors request revisions, the revised manuscript might go back to reviewers for evaluation. The length of this process very much depends on the people who are involved along the way, the promptness of their responses, and the quality of their reviews.

I therefore implore readers to be responsible members of the scientific barter economy. If you expect your peers to review your manuscripts, then you must reciprocate by reviewing theirs. I often hear people say that they are overbooked and do not have time to review manuscripts. If this is true, they should perhaps reconsider submission of their own papers to peer-reviewed journals, because in doing so they are relying on their peers to be less busy (unlikely) or to act responsibly and sustain the barter system despite their already heavy workloads. Some reluctant individuals might argue that they do not have any obligation to respond or be diligent since journals do not pay for reviews. But it is not the journals that are impacted by this stance — it is our community.

Here I present five rules for journal reviewing that will help anyone—not just new researchers—add value to the mathematical research enterprise and the associated community:

*Thousands of reviewers contribute to SIAM journals, all of which serve as high-quality venues for the publication of mathematical research. Figure courtesy of SIAM.*

**Rule 1:** Treat other writers' manuscripts the way that you would want them to treat your own.

**Rule 2:** Always be prompt when responding to inquiries from editors and editorial staff. Do not ignore a request to review a manuscript. If you plan to decline the invitation, do so straightaway; otherwise you are keeping editors from inviting other reviewers and unnecessarily delaying manuscript processing.

**Rule 3:** Make every effort to accept invitations to review if there is an opportunity for constructive commentary. This does not mean that the manuscript must be exactly within your area of expertise. However, you should be able to provide some constructive feedback to the author and editor, if only in the form of suggestions to make the manuscript more accessible to researchers within a related but distinct field. Let the editor know about your relevant proficiencies if you accept. You might even learn something new and unexpected by reading outside of your immediate research area.

**Rule 4:** Provide a review that is helpful to both the author and the editor. Odds are that the manuscript is not exactly in the editor's direct area of expertise. If your report states that the manuscript should be accepted (or rejected) without offering substantial reasons, the editor will likely ignore it and the author will be frustrated by the lack of constructive feedback.

**Rule 5:** If any issues arise, communicate them promptly with the editor/editorial staff. Such issues include delays that prevent you from completing your review assignment on time, unexpected problems with the manuscript like a suspicion of plagiarism, or a late realization that you have a conflict of interest with the author(s).

Academic mathematical research takes a village, as the saying goes, particularly when it comes to dissemination. We are fortunate that we now have many venues through which to do so — from preprint servers like the arXiv[1] where authors can post manuscripts without restrictions to journals that provide an assessment of the results' quality and impact. As such, journals serve a role that is much like that of financial ratings agencies such as Moody's or Standard & Poor's, which rely on the community's collective expertise as well. As the means of scientific communication continuously change, journals—in one form or another—will remain important outlets that are created and maintained by and for the research community. It is up to all of us to assure that they serve the community and meet its needs.


*Reinhard Laubenbacher is director of the Laboratory for Systems Medicine and a professor in the Department of Medicine's Division of Pulmonary, Critical Care, and Sleep Medicine at the University of Florida. He serves as editor-in-chief for the Bulletin of Mathematical Biology—the journal of the Society for Mathematical Biology—which is published under contract by Springer Nature.*


[1] https://arxiv.org

# Solitary Water Waves

*Nonlinear Water Waves with Applications to Wave-Current Interactions and Tsunamis*[1] *by Adrian Constantin was published by SIAM in 2011. This book, which is part of the CBMS-NSF Regional Conference Series in Applied Mathematics, explores some of the main results and recent developments in nonlinear water waves, addresses fundamental aspects of the field, and overviews several current research topics. The following text comes from chapter five, entitled "Solitary Water Waves," and is modified slightly for clarity.*

Some two-dimensional traveling water waves have a localized profile (a hump of water that drops smoothly back to a flat water level far ahead and far behind the wave crest). For these *solitary waves*, the initial profile $\eta(X,0)=\eta_0(X)$ moves at constant speed $c>0$ and without distortion in the direction $X$ of wave propagation; at a later time $t$, the profile $\eta(X,t)$ is a translation of the initial profile by an amount $ct$ in the positive $X$ direction, $\eta(X,t)=\eta_0(X-ct)$. If one sets up a solitary wave profile of depression[2]—a possibility that is within reach for laboratory experiments—it will immediately collapse into small oscillatory waves.[3] Solitary waves were first observed by John Scott Russell in 1834 (see Figure 1). He described them in a report to the British Association for the Advancement of Science in 1844 [6]:

*"I believe I shall best introduce the phaenomenon by describing the circumstances of my own first acquaintance with it. I was observing the motion of a boat that was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped — not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation—a rounded, smooth, and well-defined heap of water—which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some 30 feet long and a foot to a foot and a half in height. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phaenomenon."*

[1] https://my.siam.org/Store/Product/viewproduct/?ProductId=21535304

[2] The wave profile would be the reflection in a horizontal line of the profile of a wave of elevation.

[3] Solitary waves of depression are only possible if one allows for capillary effects; consequently, these are waves of very small amplitude. We consider gravity waves (of elevation) that propagate at the free surface of a layer of water with an average depth that measures from tens of centimeters to hundreds of meters. Displacements of the free surface are of interest from the point of view of waves at sea or in a canal.

The ability of this water wave to retain its shape for a long period of time is quite remarkable. Russell conducted numerous detailed experiments to investigate the nature of what he called "the great wave of translation," but which came to be known as the solitary wave.

Solitary waves can be easily obtained experimentally by dropping a weight at one end of a long rectangular tank (see Figure 2). Russell was the first to observe that an initial elevation might—depending on the relation between its height and length—evolve into a pure solitary wave of elevation, a single solitary wave of elevation with some residual oscillatory waves, or two or more solitary waves of elevation with or without residual oscillatory waves. He also attempted to produce solitary waves of depression, but found that an initial depression is transformed into oscillatory waves of gradually increasing length and decreasing amplitude. From his laboratory instruments, Russell concluded that the speed $c$ of a solitary wave is given by

**FROM THE SIAM BOOKSHELF**

$$c=\sqrt{g(d+a)}, \qquad (1)$$

where $d$ is the average depth of the water above a flat bed, $a$ is the wave amplitude, and $g$ is the acceleration due to gravity. Notice that (1) indicates that higher solitary waves travel faster.

In the early days, the existence of solitary waves excited some controversy because Russell's conclusions contradicted the predictions of George Biddell Airy [1], who was considered at that moment to be the leading expert in hydrodynamics.[4] The conflict between Russell's observations and the accepted theory was resolved independently by Joseph Valentin Boussinesq in 1871 [2] and Lord Rayleigh in 1876 [5]. Their insight was to incorporate weak nonlinear effects, particularly an appropriate allowance for vertical acceleration that was neglected in Airy's linear theory. The two showed that suitable approximations to the governing equations for water waves lead, for wave amplitudes $a>0$ that are sufficiently small, to the solitary wave solution

$$\eta(X,t)=\frac{a}{\cosh^2[\beta(X-ct)]}, \qquad (2)$$

with the wave speed $c$ given by Russell's formula (1) and $\beta=\frac{\sqrt{3a}}{2d\sqrt{d+a}}$. They did not, however, derive a model equation (via an approximation procedure that starts from the governing equations for water waves) that admits the solution (2). This final step was completed by Diederik Korteweg and Gustav de Vries in 1895 [4]. For present purposes, we consider not the equation written in physical variables, but its normalized form

[4] More details on these fascinating historical aspects are available in [3]. Basically, the linear theory of waves of small amplitude fails to yield any approximation to solitary waves [7].

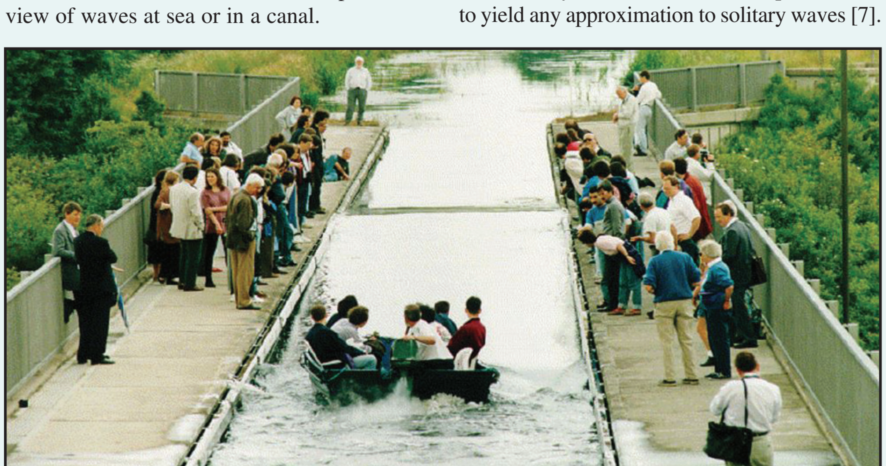


**Figure 1.** *On July 12, 1995, an international gathering of scientists witnessed a re-creation of John Scott Russell's famous "first" sighting of a solitary wave on the Union Canal near Edinburgh, Scotland. Photo reproduced with permission from the Department of Mathematics at Heriot-Watt University in Edinburgh.*

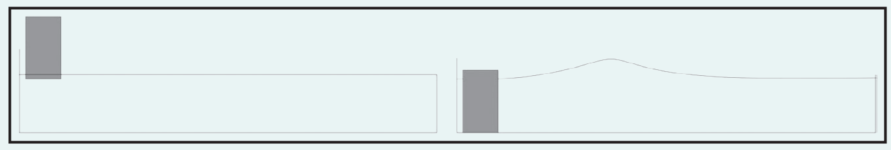


**Figure 2.** *Sketch of John Scott Russell's experiment.*

$$\eta_t+\eta\eta_X+\eta_{XXX}=0. \qquad (3)$$

We look for a solitary wave solution $\eta(X,t)=f(X-ct)$, where $c>0$ and $f(s)$, $f'(s)$, and $f''(s)$ tend to zero as $|s|\to\infty$. One can integrate (3) by substituting this form, but it is easier to check that for any constant $\alpha\in\mathbb{R}$, the wave profile

$$\eta_\alpha(X,t)=\frac{3c}{\cosh^2(\frac{\sqrt{c}}{2}(X-ct)+\alpha)} \qquad (4)$$

represents a solitary wave solution of the Korteweg-de Vries (KdV) equation.[5] At any fixed time $t$, the graph of the solitary wave (4) is formed by a hump that drops smoothly and rapidly to zero away from its crest, which is located at $X=ct-\frac{2\alpha}{\sqrt{c}}$. Russell observed in his experiments that solitary waves with greater amplitude—defined for a solitary wave of elevation as the maximal elevation of the wave profile above the asymptotic flat level[6]—travel faster and are narrower. This is borne out in the solution of (4), as the amplitude $a$ of the wave is three times the wave speed $c$, and the width of the wave—defined as the distance between the points of height $a/2$—is inversely proportional to the square root of its amplitude. Note the similarity between (4) and (2).

[5] It is not difficult to check that these are all of the solitary wave solutions.

[6] Notice that this level is beneath the mean water level $Y=\int_{\mathbb{R}}\eta(x)dx$ — in the absence of elevation, the amount of water raised above it cannot disappear in view of mass conservation.

## References


[1] Airy, G. (1841). Tides and waves. *Encyclop. Metrop.*, *3*.

[2] Boussinesq, J.V. (1871). Théorie de l'intumescence appelée 'onde solitaire' ou 'de translation', se propageant dans un canal rectangulaire. *C.R. Acad. Sci. Paris*, *72*, 755-759.

[3] Craik, A.D.D. (2004). The origins of water wave theory. *Annu. Rev. Fluid Mech.*, *36*, 1-28.

[4] Korteweg, D.J., & de Vries, G. (1895). On the change of form of long waves advancing in a rectangular canal and on a new type of long stationary waves. *Philos. Mag.*, *39*(240), 422-443.

[5] Rayleigh, B. (1876). On waves. *Philos. Mag.*, *1*(4), 257-279.

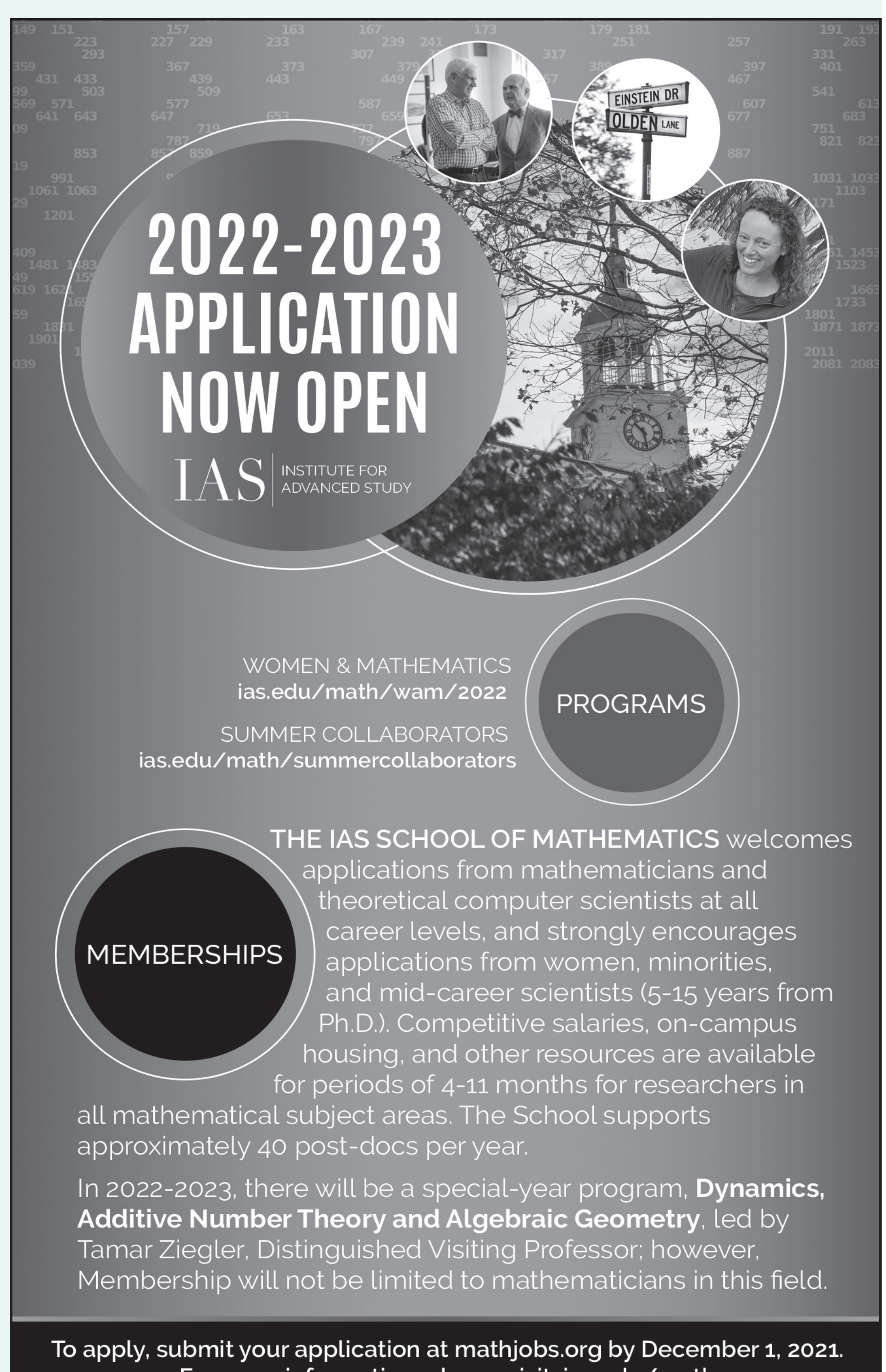
[6] Russell, J.S. (1844). Report on waves. In *Report of the fourteenth meeting of the British Association for the Advancement of Science*. York, U.K.

[7] Stoker, J.J. (1957). *Water waves: The mathematical theory with applications*. New York, NY: Interscience Publishers.


*Enjoy this passage? Visit the SIAM Bookstore*[7] *to learn more about Nonlinear Water Waves with Applications to Wave-Current Interactions and Tsunamis and browse other SIAM titles.*


*Adrian Constantin is a professor at the University of Vienna who does research in the field of partial differential equations, especially on aspects related to wave propagation.*


[7] https://my.siam.org/Store

## NIST Big Data

*Continued from page 8*

NBDIF provides a well-defined interface specification to support the application programming interfaces between NBDRA components and provision big data systems with hybrid and multiple frameworks. The service specification is based on OpenAPI,[5] a resource-oriented architecture that provides a general form of interfaces that are independent of the framework or computer language in which the services are specified. As a result, NBDRA can enable instantiation to support multiple big data architectures across different application domains.

### Big Data Security and Privacy

The security and privacy fabric is a cross-cutting, fundamental aspect of NBDRA. It represents the intricate, interconnected nature and ubiquity of security and privacy throughout NBDRA's main components. Common security and privacy concerns for all items include authentication, access rights, traceability, accountability, threat and vulnerability management, and risk management.

Additional security and privacy considerations exist at the interface levels between NBDRA components, which include the following components:

– Interface between the data provider and big data application provider

– Interface between the big data application provider and data consumer

– Interface between the big data application provider and big data framework provider

– Internal interface within the big data framework provider

– Inside system orchestrator.

Security and privacy measures are becoming increasingly critical with the escalation of big data generation and utilization via public cloud infrastructure that is built with the help of various hardware, operating systems, and analytical software. The surge in streaming cloud technology necessitates extremely rapid responses to security issues and threats [1].

[5] https://www.openapis.org

### Big Data Standards Roadmap

NIST's Big Data Standards Roadmap provides a vision for the future of big data. It assesses the current technology landscape and identifies available standards to support present and future big data domain needs from an architecture and functionality perspective. NBDIF has established a set of criteria that evaluates whether standards meet the NBDRA general requirements:

– Facilitate interfaces between NBDRA components

– Facilitate the handling of data with one or more big data characteristics

– Represent a fundamental function that one or more NBDRA components must implement.

As most standards represent some form of interface between components, one can identify a partial collection of such standards based on the NBDRA components.

### Big Data Adoption and Modernization

The adoption of big data systems is not yet uniform throughout all industries or sectors, and some common challenges across all industries could further delay this process. A report by the U.S. Bureau of Economic Analysis and McKinsey Global Institute [3] suggests that the most apparent barrier to leveraging big data is access to the data itself. The report identifies a definite relationship between the ability to access data and the potential to capture economic value across all sectors and industries.

Strategies to modernize big data may include cost estimation, workforce allocation, project prioritization, technology selection, transition plans, road mapping, and implementation timelines. In addition, those who are making governance decisions about technology should pay special attention to business models and data requirements. The resulting resolutions can help inform future strategies.

| NBDIF Volumes | ISO/IEC Documents (publication date) |
|---|---|
| Vol. 1: Big Data Definitions (Version 1, September 2015) | 20546 Big Data Overview and Vocabulary (February 2019) |
| N/A | 20547-1 Big Data Framework and Application Process (April 2020) |
| Vol. 2: Big Data Taxonomies (Version 1, September 2015) | N/A – part of 20547-3 |
| Vol. 3: Big Data Use Cases and Requirements (Version 1, September 2015) | 20547-2 Big Data Use Cases and Derived Requirements (April 2018) |
| Vol. 4: Big Data Security and Privacy (Version 1, September 2015) | 20547-4 Big Data Security and Privacy (October 2020) |
| Vol. 5: Big Data Architecture Survey White Paper (Version 1, September 2015) | N/A |
| Vol. 6: Big Data Reference Architecture (Version 1, September 2015) | 20547-3 Big Data Reference Architecture (March 2020) |
| Vol. 7: Big Data Standards Roadmap (Version 1, September 2015) | 20547-5 Big Data Standards Roadmap (April 2018) |
| Vol. 8: Big Data Reference Architecture Interfaces (new, October 2019) | *Submitted March 2020, under consideration* |
| Vol. 9: Big Data Adoption and Modernization (new, October 2019) | *Submitted March 2020, under consideration* |

**Figure 3.** *Relationship between NIST Big Data Interoperability Framework (NBDIF) volumes and documents for the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC). Figure courtesy of NIST Big Data Public Working Group.*

### NBDIF as an International Standard

In late 2015, NIST submitted the initial version of seven volumes of NBDIF to the Joint Technical Committee of the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC JTC 1) for standards development consideration. More than 200 experts from over 25 countries reviewed and contributed to the content. As a result, five of the seven original NBDIF volumes are now part of the ISO/IEC standards. Figure 3 depicts the mapping between NBDIF volumes and ISO/IEC documents.

### NBD-PWG Next Steps

With big data's compound annual growth rate at 61 percent and its ever-increasing deluge of information, the collective sum of world data is projected to grow from 33 zettabytes (ZB, $10^{21}$) in 2018 to 175 ZB by 2025 [2]. Such a rich source of information requires a massive analysis that can effectively bring about much insight and knowledge discovery.

NBD-PWG is exploring how to best extend NBDIF for the packaging of scalable analytics as services in order to meet rapid information growth. Regardless of the underlying computing environment, these services would be reusable, deployable, and operational for big data analytics, high-performance computing, and artificial intelligence machine learning applications.

### References


[1] Cloud Security Alliance. (2013, April). *Expanded top ten big data security and privacy challenges* (Big Data Working Group). Retrieved from https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf.

[2] Coughlin, Tom. (2018, November 27). 175 Zettabytes by 2025. *Forbes*. Retrieved from https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/#587370055459.

[3] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011, May 1). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation.


*Wo Chang is the digital data advisor for the National Institute of Standards and Technology (NIST) Information Technology Laboratory. His responsibilities include promoting a vital and growing big data community at NIST with external stakeholders in commercial, academic, and governmental sectors.*


CALL FOR
SIAM Prize NOMINATIONS

2022 Prizes

AWM-SIAM Sonia Kovalevsky Lecture

George Pólya Prize in Mathematics

John von Neumann Prize

Julian Cole Lectureship

Richard C. DiPrima Prize

SIAM Student Paper Prizes

W. T. and Idalia Reid Prize

Why Nominate Someone?

- Recognize and reward outstanding accomplishments
- Bring honor and prestige to your department or place of work
- Demonstrate the importance of your field to students, research funders, and the scientific community at large

The SIAM Prize Program provides an opportunity to honor members of our community — from lifetime contributions to student and early career recognition.

SUBMIT YOUR NOMINATION BY OCTOBER 15!
siam.org/deadline-calendar

siam PRIZE PROGRAM

SIAM CAREER FAIR

Tuesday, November 9, 2021
11 a.m.–2 p.m.
3 p.m.–5 p.m. EDT
HAPPENING VIRTUALLY

More details about the event will be posted at:
*go.siam.org/careerfair2021*
Think your institution, lab, or company would like to participate?
Contact marketing@siam.org

siam

# AN21 Panelists Offer Publication Advice to Prospective Book Authors

*By Lina Sorg*

Writing and publishing one's own book may seem like a daunting task, especially for first-time authors. However, certain strategies and approaches can simplify and facilitate the process. During the 2021 SIAM Annual Meeting,[1] which took place virtually in July, a panel of SIAM book authors engaged in a lively discussion about the nuances of book publication. The panel—which was chaired by Elizabeth Greenspan of SIAM's Book Acquisitions division—included Emily Evans[2] (Brigham Young University), Nicholas Higham[3] (University of Manchester), Ralph Smith[4] (North Carolina State University), and Paula Callaghan (SIAM). All of the panelists reflected on their respective techniques, offered guidance for anyone who is thinking about writing a book, and spoke about the advantages of publishing with SIAM.

The panelists began by urging potential authors to think deeply about the motivations and desires that inspire their prospective projects. For example, one might wish to address new teaching methods, novel approaches to existing research topics, or unique software developments. Higham encouraged attendees to consider whether they would use their imagined books themselves. "I've written four SIAM books and I use them all quite regularly," he said, adding that authors who do not plan on consulting their own books might not have found the right premise or angle.

Conversation quickly moved to the topic of book proposals, which reflect a book's aim and scope and identify the way in which it differs from existing literature. Some authors choose to submit proposals early in the writing process, but Smith opts to wait until he has drafted roughly 75 percent of the text. "It takes me until that point to see most of the final picture of the book," he said. Greenspan noted that the amount of manuscript that authors submit to SIAM varies, ranging from a table of contents and preface to a complete draft. More senior individuals who have long records of publication might submit only the former, but junior researchers should submit at least one chapter. "We do get a lot of proposals where the book is half done or even sometimes more," Greenspan said. "Sometimes people just don't want the pressure of signing a contract without having written most of the book."

Before submitting a proposal, Smith recommended that textbook writers teach a course with a draft of their textbook to gauge student reactions and ensure that the content meets its targeted audience — he even allows students to edit and offer suggestions about the material. In fact, virtually all SIAM textbooks begin as lecture notes. When translating course notes into book format, authors must keep the knowledge level of their intended audience in mind.

Like Smith, Evans touted the value of teaching-based feedback. She gave her drafted textbook to a colleague, asked him to teach from it, and solicited candid comments based on his experience. She also had additional coworkers and graduate students review the text, incentivizing the latter with extra credit for each mistake that they found. "Students are much more likely to express their opinions if they're getting something out of it," Evans said.

While authors often ask their students or colleagues to look over proposal materials, a fundamental part of the proposal process for both textbooks and monographs is the outside reviews that the publisher gathers after proposal receipt. Unlike commentary from students or colleagues, these reviews—as with journal articles—are anonymous and come from individuals who are experts in the field but have no personal involvement in the project at hand. This feedback can also jumpstart the writing process for authors who are suffering from writer's block.

As such, employing multiple reviewers is beneficial. "Send as many people through the book as possible," Smith said. "It's amazing how different people look at it with different lenses." He noted that colleagues might have ideas for better approaches and students are particularly quick to identify concepts or explanations that they do not understand. However, he warned that the review process can be tedious when one reaches out to multiple critics. For instance, he once spent an entire month revising one section based on a single comment from a SIAM reviewer.

Evans remarked that in some cases, reviewers should do more than simply read the text. "When you're publishing a book, you need to have someone—and it doesn't have to be the same person—do every single one of the exercises," she said. This guideline becomes even more important in the context of programming, as a third party should run and test all software.

*A panel at the 2021 SIAM Annual Meeting, which took place virtually in July, offered strategies and approaches to facilitate the process of writing and publishing a book. Top row, left to right: Paula Callaghan (SIAM), Elizabeth Greenspan (SIAM), and Emily Evans (Brigham Young University). Bottom row, left to right: Ralph Smith (North Carolina State University) and Nicholas Higham (University of Manchester).*

Discussion then turned to the perks and drawbacks of working with coauthors. Higham, who has written books by himself and with other people, commented that each path has its own benefits. Some writers choose to recruit coauthors because they cannot complete the prospective book themselves and need additional expertise. For sizable projects, sharing the work may mean that the final product publishes in a more timely manner. Nevertheless, Higham suggested that coauthors first write a journal paper together to ensure that their work styles are compatible. "Do think very carefully before you start a coauthor project," he said. "The dangers of coauthorship are that it might slow things down."

Evans, who wrote *Foundations of Applied Mathematics, Volume 1: Mathematical Analysis* with two coauthors, conceded the risks but offered a different perspective. She found that having coauthors motivated her when she did not feel like writing because it held her accountable to other people. Collaboration also provides an opportunity for writers to enlist individuals with unique capabilities, such as pedagogical proficiency. "If you're picking coauthors, you need to pick people who have strengths or expertise that very much complement your own," Evans said.

Regardless of the number of authors, the panelists agreed that writing a book always takes longer than expected. "In all three cases I had an initial schedule, and in all three cases I didn't quite finish when I said I was going to," Smith said. "One of the things I could caution all authors is that the last 15 to 20 percent of the book seems to take 40 percent of the time." If writers are testing their material by teaching it in a course, they can link their writing schedules with their semester syllabi for pacing purposes. Smith also advised attendees to allow sufficient time for the incorporation of feedback.

Though authors should of course attempt to meet their deadlines, Higham indicated that many publishers would rather have writers submit their books a few months late than compromise the quality by rushing to finish. "Don't agree to a deadline unless you're really happy with it," he said. "You might need more time to put the book aside." Greenspan noted that people who submit proposals and contracts in a book's early stages are more likely to miss deadlines, while those who deliver proposals for nearly-written books tend to finish much closer to the agreed-upon target.

All of the panelists conceded that preparing the index is one of the most time-consuming parts of the process because indices take a long time to both create and debug. Higham urged writers to refrain from saving the index until the end, as working on it several months before the book's completion allows one to catch LaTeX errors and other inconsistencies. Smith concurred and recommended that authors allot specific blocks of time exclusively for indices, which are important tools for readers. "For a general reader, the index is really important," he said. "Please take the index very seriously."

After successfully publishing a book, authors might eventually wonder whether they should consider a second (or third, or fourth) edition. SIAM recommends that at least 20 to 25 percent of material in a subsequent edition be new and different from the existing text. This material could stem from a dramatic breakthrough, a novel technique, or a new software. In some cases, future editions of certain books—such as those that address software-oriented topics like MATLAB—are guided by changes in the software itself. "There clearly has to be something different that makes the new edition warranted," Higham said. "In particular, why would people want to buy the new edition rather than the old edition?" For example, one need not immediately consider a second edition if the field in question is not rapidly changing.

Higham keeps files for all his books and makes annotations about items that he spots after publication—including typos, references, and other updates—that he would like to incorporate or improve upon in future editions. This system allows him to maintain organized notes about his thought processes and helps him prepare for any subsequent publications. In some cases, authors can make minor typo changes and other corrections if/when a book enters a second printing cycle. "SIAM is very good at working with you on that, and that might not universally be the case," Smith said.

As the panel wrapped up, panelists reflected on their publishing experiences with SIAM. Higham praised SIAM copyeditors who correct grammar, spelling, references, quotations, and sentence structure, and added that he cannot imagine publishing with an organization that does not provide copyediting services. "Copyeditors are extremely important," he said. "They really do improve my best efforts. The final result is a book that is far better from what I delivered."

Higham concluded with several remarks about the consistency and flexibility of SIAM publishing. He noted that most authors work with the same staff members throughout the entire publication process, from proposal submission to final edits (sometimes a few years later). "You get to know people and they'll take you through the whole process, beginning to end," he said. "If you have a reasonable request about your book, SIAM will always listen and accede to your request if possible."

*Lina Sorg is the managing editor of SIAM News.*

[1] https://www.siam.org/conferences/cm/conference/an21

[2] Co-author of *Foundations of Applied Mathematics, Volume 1: Mathematical Analysis*

[3] Author/co-author of *Accuracy and Stability of Numerical Algorithms* (2nd edition), *Functions of Matrices: Theory and Computation, MATLAB Guide* (3rd edition), and *Handbook of Writing for the Mathematical Sciences* (3rd edition)

[4] Author of *Smart Material Systems: Model Development* and *Uncertainty Quantification: Theory, Implementation, and Applications*