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

siam news

Volume 50/ Issue 4
May 2017

Special Issue on Computational Science and Engineering

Check out articles from the 2017 SIAM Conference on Computational Science and Engineering—and more—in this month's ***special issue****!*

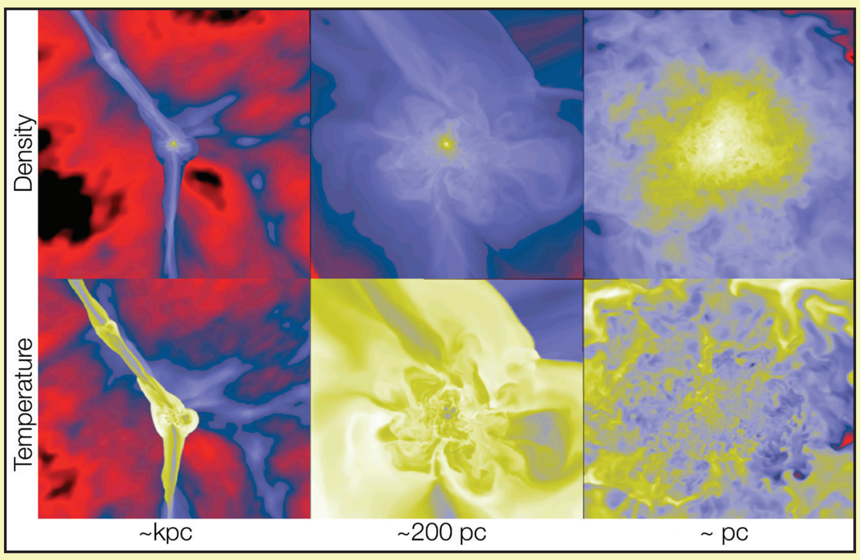


Figure 1. *One of the successes of astrophysical computation via adaptive methods: isodensity and isothermal profiles at three levels of spatial resolution during the evolution of the very first stars. "Pc" abbreviates parsec, an astronomical unit of length equivalent to about 3.25 light-years, or 31 × 10^{12} km. Image credit: Matthew Turk and Tom Abel (KIPAC/Stanford).*

In an article on page 6, Paul Davis reports on various applications of adaptive mesh refinement and looks towards the next generation of the method's use.

CSE Achievements in Aircraft Design

By Paul Davis

The aeronautics industry is a sophisticated user of computational science and engineering (CSE), and a source of some of the field's most difficult challenges. Bruno Stoufflet, Chief Technology Officer at Dassault Aviation and a recognized CSE expert, brought that message home with an insider's detailed account of the field's expanding role to attendees of the 2017 SIAM Conference on Computational Science and Engineering (CSE17), held this February in Atlanta, Ga.

Dassault Aviation designs and manufactures business jets, military fighters, and unmanned aerial vehicles (commonly called drones). It develops most of its simulation codes in-house, with assistance from external collaborators. Embodying the firm's wide scientific awareness, Stoufflet cited the value of the finite element work of Thomas Hughes, who was in the audience and had just received the SIAM/ACM Prize in Computational Science and Engineering (see page 4).

At Dassault, CSE plays a role in three phases of an aircraft's life cycle: design, development, and post-delivery support. Design commands by far the largest share of Dassault's CSE efforts, most of which promote a traditional engineering perspective (although newer stochastic approaches are emerging). Stoufflet's long list of design-centered challenges included industrial-scale computational fluid dynamics (CFD), automatic shape optimization, multi-physics analyses, computational electrodynamics, surrogate modeling, uncertainty quantification for robust design, and the movement of all these computational tasks to exascale environments.

A typical computational task during the development phase is evaluation of the probability of rare events, such as a collision during the release of a store from an aircraft. While some of the variables are under the pilot's control, others are not. Brute-force estimates are not feasible, so importance sampling is the primary tool. In a somewhat different direction, computation's role in support following delivery typically involves data analytics that add value by justifying reliability estimates or guiding predictive maintenance, for example.

Stoufflet emphasized the multidisciplinary character of the design loop, which consumes the lion's share of Dassault's

See **Aircraft Design** *on page 3*

Cotranscriptional Folding: A Frontier in Molecular Engineering

A Challenge for Computer Scientists

By Shinnosuke Seki

Transcription is a process that synthesizes a temporal copy of a gene—called a *transcript*—out of an RNA molecule. The transcript acts as an intermediary to express a protein or non-coding RNA encoded in the gene. A gene is a chemically-directed chain of deoxyribonucleotides A, C, G, and T. The synthesis proceeds sequentially. An enzyme called *RNA polymerase* scans a gene unidirectionally and binds one ribonucleotide of the type (A, C, G, or U) most energetically preferred by what is read to the growing transcript; the preference is $A \to U$, $C \to G$, $G \to C$, and $T \to A$. Upon further modification, the transcript may adopt a precise tertiary structure (*conformation*), thus allowing it to perform its function.

Some modifications, such as the removal of introns (splicing) from a transcript, are cotranscriptional: as transcription occurs, introns fold into a loop and are excised. *Cotranscriptional folding* occurs when a transcript folds upon itself during synthesis. Such folding is possible because transcripts fold considerably faster than they synthesize. The relative speed of folding to synthesis seems to be predetermined in nature, but the reason for this is unknown. Indeed, computation of thermodynamics indicates that polymerase deceleration could save energy [3]. The specific *transcription rate* and directionality of synthesis enable nature to "program" a gene's biological function [8], with polymerase as a compiler from the genetic "source code" back to the executable program, that is, biological function. The source code offers inheritability to nature and readability to us.

In 2014, Cody Geary, Paul Rothemund, and Ebbe Andersen exhibited a command of this programming language as *RNA origami* [5], fabricating complicated shapes—such as rectangular tiles—from RNA molecules (see Figure 1). The tile hinges its left and right halves—each of which is a stratus of hairpins, i.e., strong double-helical stems ending in a loop—at the bottom. First, the left half cotranscriptionally folds through a pathway of events to form the hairpins, then the right half folds as such and is stapled cotranscriptionally to the folded left half. The two halves are held together weakly via paranemic "kissing" loops (see Figure 2, on page 4). By the time the right half begins undergoing synthesis, the left has already been constrained too strictly to form a strong double-helical segment with the right.

RNA origami will likely be the "Hello, World!" of future educational materials on the programming of cotranscriptional folding: a symbolic and static code that always behaves consistently. Readers will then move to chapters on more dynamic codes for information processing. The *oritatami system* is a mathematical model recently proposed by Geary, Pierre-Étienne Meunier, Nicolas Schabanel, and myself to establish theoretical grounds for these chapters [4]. As shown in Figure 2 (on page 4), the oritatami system abstracts a conformation as a directed, vertex-labeled path (transcript) on the triangular grid with bonds between adjacent vertices. Vertices on the path are called *beads*. A bead may represent one nucleotide or a region of consecutive nucleotides depending on the abstraction level, and its label is taken from an alphabet Σ of bead types. For *x*, $y \in \Sigma$, an *x*-bead can bind with a *y*-bead only if the pair (x, y) belongs to a *rule set* $\mathcal{H}$, a symmetric relation in Σ^2. Two other parameters exist: the *delay* δ abstracts the transcription rate and the *arity* α bounds the number of bonds that one bead can form. An oritatami system is a tuple $(\Sigma, w, \mathcal{H}, \delta, \alpha, \sigma)$. Upon its initial "seed" conformation σ, the system transcribes the first δ beads of its transcript $w \in \Sigma^*$ and repeats the following instructions until the end of w:

1. Fold the fragment of nascent beads to elongate the current conformation with as many new bonds as possible.

2. Stabilize the eldest nascent bead with all its bonds accordingly.

3. Transcribe the next bead, if any.

See **Cotranscriptional Folding** *on page 4*

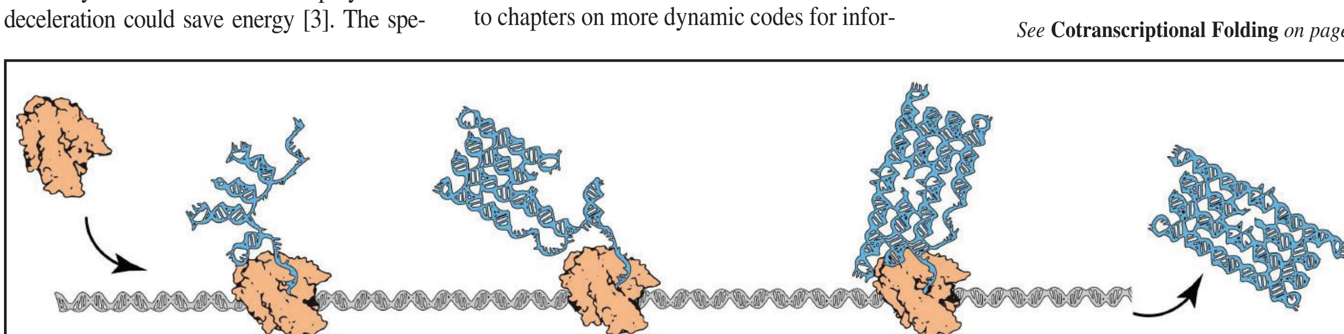


Figure 1. *RNA origami. The transcript (blue) folds into a rectangular tile while being synthesized by the RNA polymerase (orange). Image courtesy of [5].*

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 50/ Issue 4/ May 2017

In Search of the Perfect Numerical Analysis Textbook

FROM THE SIAM PRESIDENT
By Nicholas Higham

My bookshelf contains a lot of numerical analysis textbooks. The oldest is Douglas Hartree's *Numerical Analysis* (Oxford University Press, 1958), and the newest is Robert Corless and Nicolas Fillion's *A Graduate Introduction to Numerical Methods: From the Viewpoint of Backward Error Analysis* (Springer, 2013).[1]

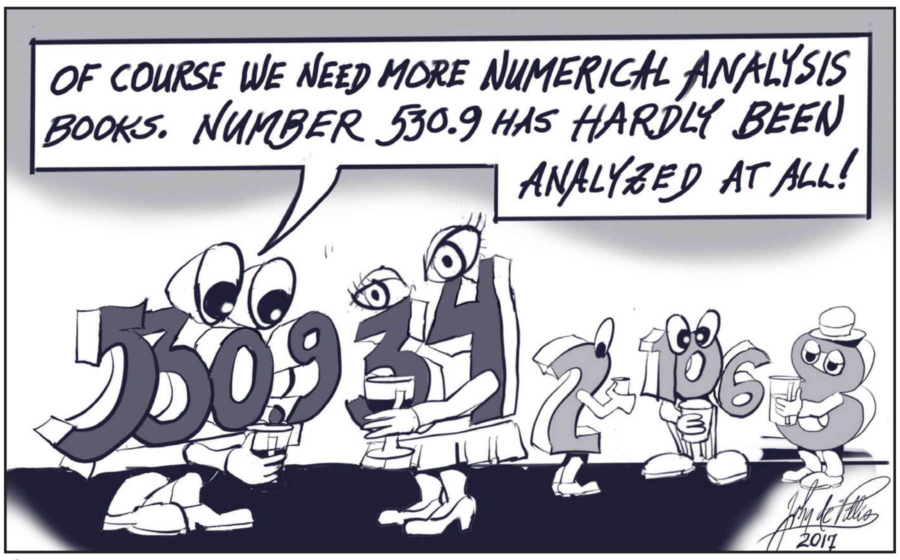
The variety is wondrous, ranging from books at an introductory level to those aimed at an advanced graduate audience; from formal treatments with numerous theorems to more computationally-oriented presentations; and from books tied to a particular programming language to those that are language-independent. Why do I keep acquiring them? Why do authors keep writing them?

I am not the first to ask the latter question in *SIAM News*. In a lighthearted 1984 article titled "On Therapy and Numerical Analysis Texts," Paul Davis called writing a numerical analysis textbook "the leading case of overwork among academic mathematicians," and asked, "Why are we driven to this insanity?"

These questions are certainly relevant since SIAM publishes many textbooks in numerical analysis — both general and in specific areas of the subject. Many of us have at some time struggled to find a completely satisfying textbook for a course we need to teach, an experience that seems particularly common among numerical analysts. Given that there is a fairly standard body of core material on the subject, why should we have this problem? I can see several reasons.

First, numerical analysis continues to evolve. The introduction of the IEEE arithmetic standard in 1985 made it easier to describe floating point arithmetic (removing the requirement to discuss guard digits, for example), but also harder in that it introduced features such as NaNs and subnormal numbers that may need to be covered. Polynomial interpolation is a classic topic and it might appear that nothing has changed for over half a century, but in the last decade the barycentric representation of the interpolant has become the representation of choice in many contexts, and hardly any textbooks treat it. The evolution of computer architectures may have little effect on a first course, but it certainly can influence the state of the art in more advanced courses: method A might require more flops than method B, but A could be faster if it is more parallelizable or requires less communication. Authors of books that make use of a programming language (C, Python, etc.) or

[1] Reviewed in the December 2016 issue of *SIAM Review*.



Cartoon created by mathematician John de Pillis.

a problem-solving environment (MATLAB, Maple, Jupyter Notebook, etc.) face a constant battle to keep up with changes in the language and software, while few, if any, textbooks use newer languages (I am not aware of any textbook that uses Julia).

A second reason for dissatisfaction with numerical analysis textbooks may be that the material does not have the right balance of theory, algorithms, and computation. For example, on the topic of Runge-Kutta methods, should a general class of methods be derived, or one particular method stated? What types of error and stability should be analyzed? And to what extent should algorithmic practicalities be discussed?

Another reason is that the field of numerical analysis is big enough that no book can treat all the topics covered in courses: divided differences, multidimensional interpolation and integration, stationary iterative methods for linear systems, and stiff ordinary differential equations are not found in every book. And with the growing importance of stochastic computation and uncertainty quantification, there is an argument for including some relevant aspects of probability and statistics.

Application areas influence the examples included in a textbook. While the computation of PageRank is now commonly presented as a practical use of the power method, future textbooks may emphasize the relevance of the subject to machine learning, or some area that is as yet in its infancy.

As well as the choice of topics, there is no agreement on the order in which to present them. The first chapter of a numerical analysis textbook has traditionally been about errors and floating point arithmetic, but some argue that this chapter should appear later because numerical analysis is not principally about floating point arithmetic.

As an instructor looking to develop a course, if you ask yourself, "Which book has the best treatment of topic X?" then you may well find that your answer has almost as many books as topics, and this feeds the desire—to which Paul Davis referred—to produce your own text. He also noted that "The drive to write yet another numerical analysis text may arise from computation's curious mix of science and art," reasoning that everyone "is anxious to share yet another insight into the art, but none dares skimp on explaining the science." Davis observed that "the best books seem to come from the people who do the best work…apparently, there is no substitute for being there."

For all of these reasons, the perfect numerical analysis textbook does not yet exist and probably never will. Authors will continue to write their own versions of what a numerical analysis textbook should be, and SIAM will continue to publish them, at least when the usual criteria—which include correctness, distinctiveness, market size, and lack of duplication of existing SIAM books—are met.

If you have an idea for a new textbook or research monograph—in numerical analysis or any subject that fits SIAM's purview—please contact the SIAM acquisition editors,[2] who will be happy to discuss the idea with you.

Nicholas Higham is the Richardson Professor of Applied Mathematics at the University of Manchester. He is the current president of SIAM.

[2] http://www.siam.org/books/authors/proposal_info.php

Issue Acknowledgments

SIAM News would like to acknowledge Paul Davis for his notable contributions to this Special Issue on Computational Science and Engineering.

Errata and Clarifications

In the article "Explaining the East/West Asymmetry of Jet Lag," which appeared in the March Dynamical Systems Special Issue of *SIAM News*, the suprachiasmatic nucleus (SCN) was mistakenly referred to as "a tiny region of the brain's hippocampus." The SCN is part of the brain's hypothalamus, not the hippocampus.

Florin Diacu (University of Victoria) replaced Evelyn Sander (George Mason University) as the liason for the SIAM Activity Group on Dynamical Systems. We neglected to note this change in several past issues.

siam news

ISSN 1557–9573.

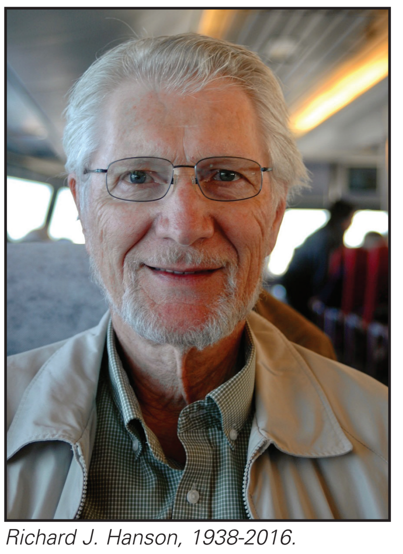
Obituaries

By Fred Krogh

Richard J. Hanson passed away on December 4, 2016, at his home in Albuquerque, N.M., after battling brain cancer for a year and a half.

Richard was born in Portland, Ore., in 1938 and received his B.S. and M.S. in mathematics from Oregon State University. We met as undergraduates at Oregon State and remained close friends since. He went on to earn his Ph.D. from the University of Wisconsin under Wolfgang Wasow, with a dissertation on turning point theory for ordinary differential equations.

Richard began his professional life as a faculty member at the University of Southern California. In the mid-1960s, he joined Charles Lawson's computational mathematics group at the Jet Propulsion Laboratory (JPL). Due to his encouragement, I joined him at JPL in 1968, only to stay on for 30 years. We were both involved with preliminary work on mathematical libraries at JPL, which led to the MATH77/mathc90 libraries.[1] Richard cowrote the first Basic Linear Algebra Subprograms (BLAS)[2] paper, and coauthored *Solving Least Squares Problems* with Lawson, published by Prentice Hall in 1974 and republished as part of SIAM's *Classics in Applied Mathematics* series in 1995.



Richard J. Hanson, 1938-2016.

After his time at JPL, Richard reentered academia with a joint position in mathematics and computer science at Washington State University. In 1976, he joined Sandia National Laboratories' Applied Mathematics Group, where he remained until 1987. During this time, Richard served as Algorithms Editor of the Association for Computing Machinery's (ACM) *Transactions on Mathematical Software* (*TOMS*) for eight years. This was followed by a short stint at Applied Dynamics International in Ann Arbor, Mich., where he developed software modules efficient enough for use in real-time-in-the-loop software.

In 1989, Richard joined the International Mathematics and Statistics Library (later Visual Numerics) in Houston, Texas. There he continued his work on numerical software. He also spent several years as a member of the research faculty in Ken Kennedy's Center for High Performance Software Research at Rice University.

Richard rejoined Visual Numerics (later acquired by Rogue Wave Software) in 2005, where he continued to work until leaving Rogue Wave in 2013. He remained professionally active up to the onset of his illness in late 2015.

Richard published numerous papers, including a few joint papers with me. He wrote almost all of his software in Fortran, keeping up with the language's standards as it evolved. This resulted in *Numerical Computing with Modern Fortran*, which Richard coauthored with Tim Hopkins. It was published by SIAM in 2013. His later work also involved issues related to taking advantage of multiple processors in mathematical software.

Richard was a longtime member of SIAM, the ACM, and the International Federation for Information Processing Working Group 2.5 on Numerical Software.

He had a lifelong enjoyment of the outdoors, and backpacked, camped, and hiked throughout the U.S. and the world. For the last ten years, Richard particularly enjoyed walking in the foothills and mountains near his home in Albuquerque. He leaves behind Karen Haskell, his wife of 39 years and enthusiastic travel companion, with whom he has also published a few papers. Richard passed along his scientific interests and abilities to his sons Eric, Joe, and Fred, and his daughter Christina. He is also survived by four talented granddaughters, two brothers, a sister, and numerous friends.

Richard was always there to listen to my strange ideas and offer good advice. And I miss him.

The author acknowledges much helpful input from Karen Haskell.

Fred Krogh (fkrogh@mathalacarte.com) is president of Math à la Carte, prior to which he worked in the Jet Propulsion Laboratory at the California Institute of Technology.

[1] http://netlib.org/math/
[2] http://www.netlib.org/blas/

Aircraft Design

Continued from page 1

CSE efforts (see Figure 1). The design loop begins with every available option for architecture and technology on the table. Analysis and optimization in the design phase are driven by such specific disciplines as aerodynamics, structures, acoustics, propulsion, and aircraft systems. Formulating parametric models enables engineers' exploration of the design space and their evaluation of sensitivities and risks. Each lap of the design loop ends with considerations of market and regulatory requirements like range, comfort, environmental restrictions, and cost.

Dassault's bread-and-butter computational fluid dynamics tools have evolved from various finite element formulations with 10,000 nodes in the early and mid-1980s to the current operational CFD code that handles upwards of 20 million grid points in less than 15 minutes on 2048-core class machines (see Figure 2). Its steady Navier-Stokes solutions and unsteady eddy simulations are used during all stages of design. "We design for cruise conditions with the CFD code," Stoufflet said. "A wind tunnel is used only for intermediate and final check-out."

In the future, design teams hope to use their Reynolds-averaged Navier-Stokes steady solver to predict drag at cruising speed to within 0.5% accuracy in a typical compute time of 30 seconds. They also wish to improve models and better understand local flow physics to determine maximum lift and explore airframe acoustics more precisely. For example, predictions of noise generated by a landing gear bay are greatly influenced by the level of geometric detail included—local flow physics is at work—but a detailed computation on a 2048-core machine can require 15 days!

Developing automatic shape optimization codes requires broad collaborations with the scientific community and "strong interaction with the design team to define and model the significant pieces," Stoufflet said. The external collaborations have contributed automatic differentiation software, feasible direction sequential quadratic programming code, and a feasible arc interior point algorithm. These tools are integrated into a gradient-computation formulation that draws on partial differential equation control ideas of Jacques-Louis Lions, who also shares credit for inventing the parareal parallel-in-time algorithm.[1]

[1] Read more about the latest developments in parareal algorithms in a report of another invited lecture from CSE17, on page 5.

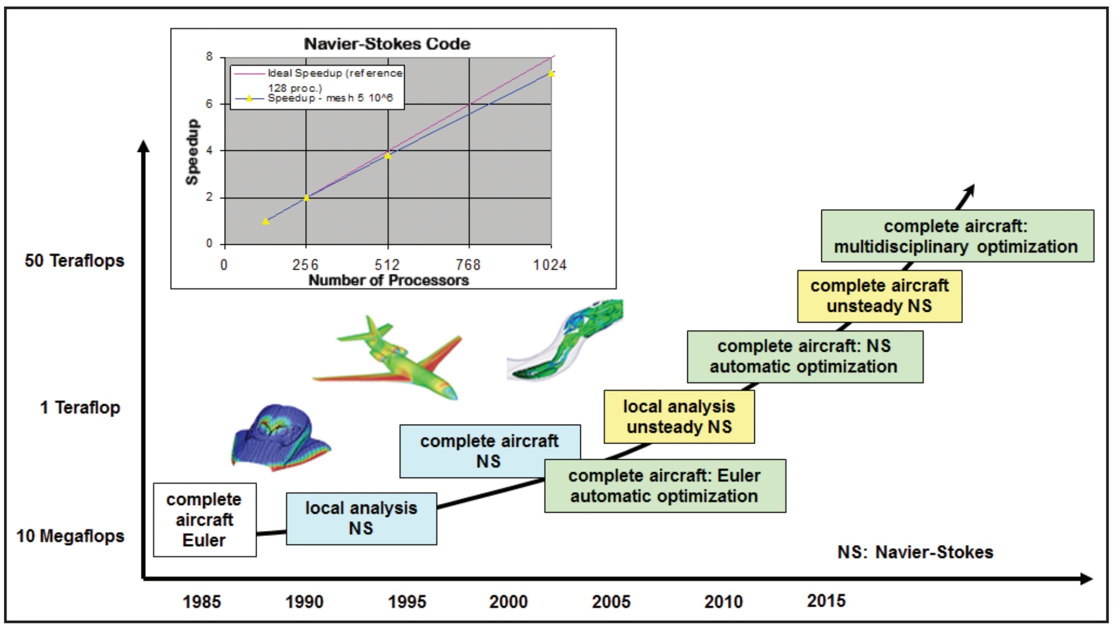


Figure 2. *The evolution over 30 years of the capabilities of computational fluid dynamics for aircraft design. Image credit: Dassault Aviation.*

When employed within the design loop, these tools have enabled a significant increase in the area of laminar flow on an aircraft wing near its fuselage, a reduction to nearly zero of an area of recirculation near an aircraft's tail, an optimum balance between low-speed lift and high-speed drag in the shape of wingtip winglets, and optimization of separated flows in curved air ducts for unmanned aerial vehicles.

Dassault's "total in-house control of tooling enables development of an optimization chain" that is in daily use, Stoufflet said. Shape optimization accelerates the early design cycle and offers wider options, while engineering analysis remains an essential step within the design loop.

Stoufflet's list of sophisticated, engineering-focused computational challenges included a multi-physics computational challenge in aero-elasticity to correctly predict the influence of a missile carried on a fighter, varied approaches to computational electromagnetics involving complex materials and geometries, and surrogate models to facilitate interactive exploration of design alternatives.

As Chief Technology Officer, Stoufflet also faces a broader, computationally-based challenge in Dassault's approach to the design process; the trade-off is robust design versus reliability-based design, or seeking to "manage uncertainties instead of adding margins." He suggests that a designer "needs a new mindset" to think in terms of the "probability of not reaching a (design) objective," rather than adding some arbitrary margins to a minimum drag requirement — for example, in hopes of accounting for the effects of possible twisting of the wing, changes in trailing edge camber, and the like.

As mindsets change, designers will need ways to propagate uncertainty through a system and access to computed response surfaces that incorporate second-derivative data, which in turn require fast CFD solvers and automatic differentiation tools.

At Dassault, Stoufflet anticipates "an unceasing effort to increase the efficiency of the design process," continued exploitation of the benefits of quantifying uncertainty, increased confidence of engineers in stochastic approaches, and the exploration of "applications of data analytics that bring added value, e.g., justifying the correctness and reliability of machine learning" algorithms.

Addressing his audience directly, Stoufflet re-emphasized the importance of the CSE community's rigorous approach while acknowledging that industry is more ad hoc. He requested that his listeners guide their own work by asking themselves, "What can we put into our framework that is rigorous *and* helps industry?"

Stoufflet's CSE17 presentation is available from SIAM either as slides with synchronized audio, or as a PDF of slides only.[2]

Paul Davis is professor emeritus of mathematical sciences at Worcester Polytechnic Institute.

[2] https://www.pathlms.com/siam/courses/4150/sections/5839

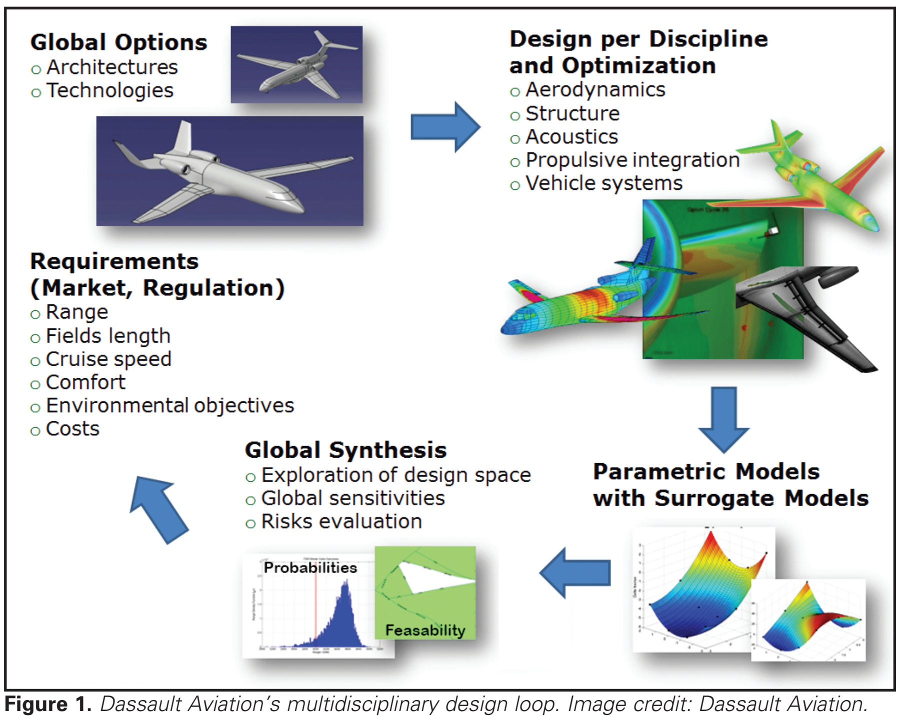


Figure 1. *Dassault Aviation's multidisciplinary design loop. Image credit: Dassault Aviation.*

Cotranscriptional Folding

Continued from page 1

The oritatami system thus folds its transcript w cotranscriptionally. Figure 3 depicts a delay-3 oritatami system folding a rigid directional structure motif nicknamed *glider*.[1] Any bead or its absence outside the circle of radius $\delta + 1$ centered at the last bead stabilized cannot affect the fragment of nascent beads. In this sense, the circle is called the *event horizon* or *context*.

In computing, it seems essential for oritatami systems to refer to some kind of "memory." Random access is one of the most essential capabilities for computation, and most computational models are either equipped with a random-access memory or capable of readily implementing it. The former is not the case for oritatami systems. Regardless of the transcript's prior folding, subsequent transcription is predetermined by w and does not change. Oritatamists have therefore speculated that the only plausible way for oritatami systems to "remember" is to equip a transcript with the capability of sensitively folding into more than one conformation to multiple possible event horizons. *Two possible event horizons and corresponding conformations may carry 1-bit information*.

This turned out to be the central principle in designing an oritatami binary counter [4]. Its transcript is periodic as (0)−(1)−(2)−···−(58)−(59)−(0)−(1)··· of period 60, where integers represent bead types. The periodically-occurring factor (0)−(1)−···−(11) along the transcript is called a *half-adder module*. The transcript folds cotranscriptionally upon the seed shown in Figure 4 (top). The event horizons that the first two occurrences of the factor encounter are distinct enough to favor different conformations, (see Figure 4, middle and bottom). The rule set of the system is designed carefully so as to equip the half-adder module with the capability of folding into four conformations, depending on the sequence of the four bead types above, which encodes a 1-bit input, and on whether a module starts folding at the top (carry = 0) or bottom (carry = 1). These four conformations expose different sequences of bead types distinguishable enough to encode a 1-bit output, and place its last bead (of type 11) at the top or bottom to represent a carry. For instance, the event horizon encountered by the first half-adder module feeds 0 with carry to the module and causes the module to fold so as to output 1 without carry, while the event horizon for the second feeds 0 without carry and folds the module so as to output 0 without carry. Two event horizons encoding 1 with or without carry fold a half-adder module into one of the other two possible conformations, respectively. The system is designed such that half-adders never encounter any other unexpected event horizon.

Due to further developments, an oritatami system to simulate a cyclic tag system is now nearly complete. The cyclic tag system is capable of simulating a Turing machine [2], and the cyclic tag system simulator would prove that the oritatami system is Turing-universal, that is, capable of computing all computable functions. Innovative features of the simulator include cyclically-accessible memory, transcription-halting machinery, geometrical "hill-and-dale" encoding, and the glider. Turing-universality marks the first key milestone, suggesting implementability of artificial inheritable stored-program computers.

With hope and the meager tools invented thus far, oritatamists venture into the frontier of cotranscriptional folding. The next milestone is *intrinsic universality*, the capability of one oritatami system to mimic the behavior of an arbitrary other. An intrinsically-universal oritatami system would offer a prototype of a universal programming language for RNA cotranscriptional folding. It seems quite distant, though, and there are many problems to be settled. The oritatami system discards some kinetic and topological details of RNA cotranscriptional folding to shed light on its computational aspects. Expertise and insights from diverse fields such as RNA kinetics, molecular dynamics, and topology would complement the oritatami system and consolidate theory and experiments into an *in-vivo* execution of basic computational steps by RNA cotranscriptional folding.

Acknowledgments: This work is in part supported by JST Program to Disseminate Tenure Tracking System No. 6F36 and JSPS KAKENHI Grant-in-Aid for Young Scientists (A) No. 16H05854. The author appreciates Cody Geary, Yo-Sub Han, Hwee Kim, Trent Rogers, Miyako Satoh, and Nicolas Schabanel for their valuable comments on earlier versions of this article.

[1] View an animation of the delay-3 oritatami system at http://www.dailymotion.com/video/x3cdj35

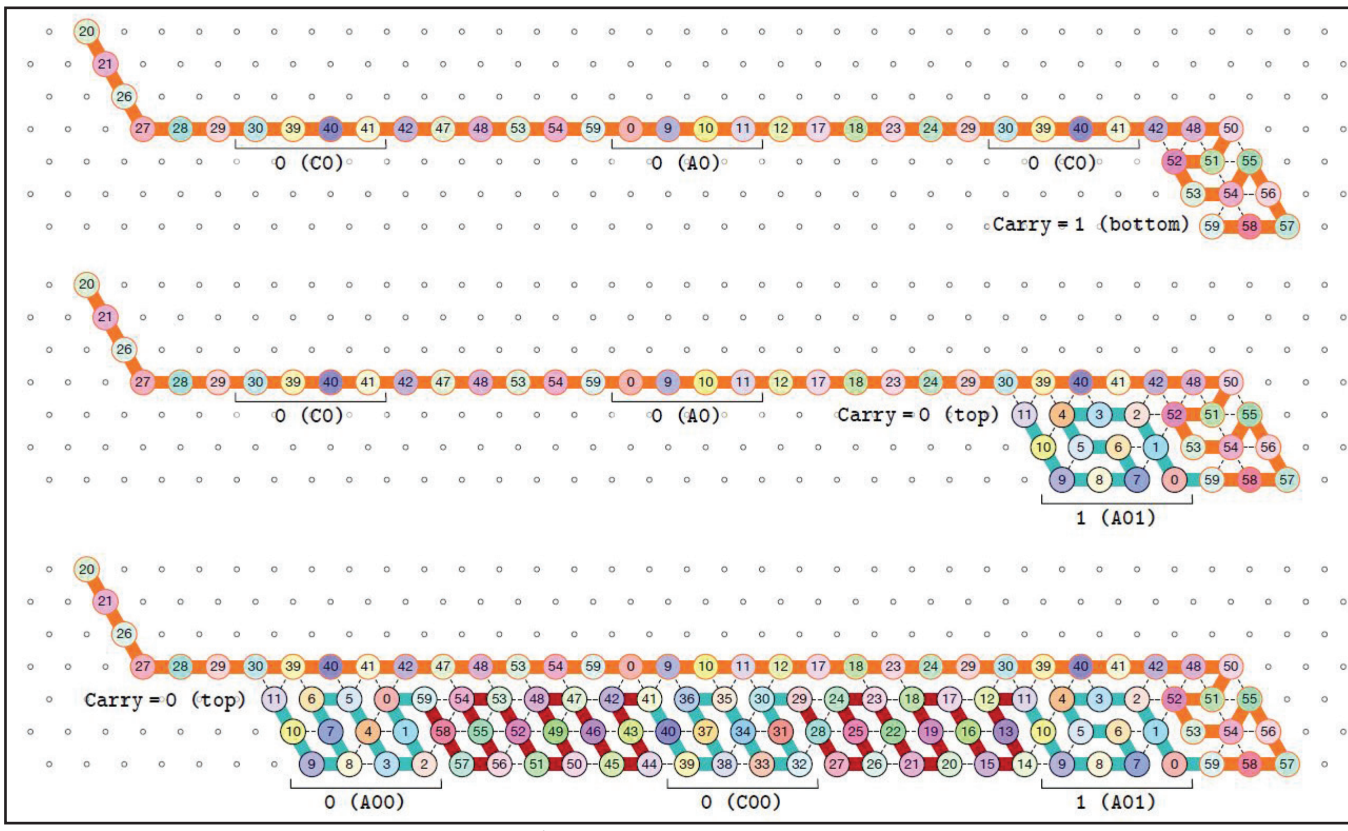


Figure 4. *An oritatami binary counter. Image courtesy of [4].*

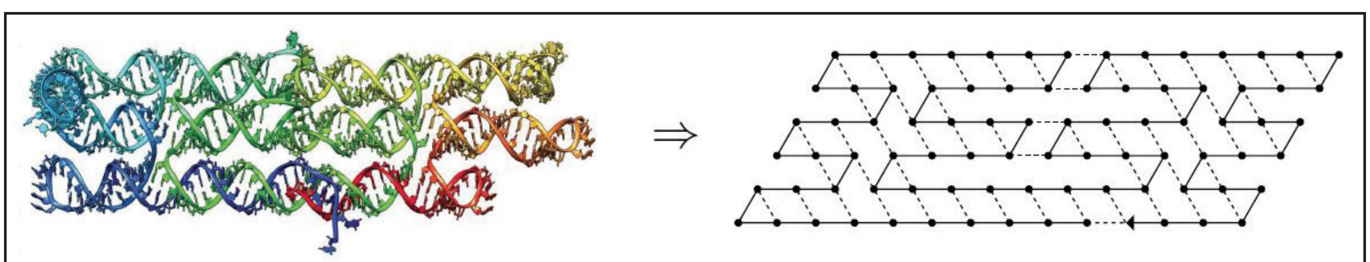


Figure 2. *Abstraction of the RNA tile structure as a conformation in an oritatami system. The dashed lines indicate bonds. Left image courtesy of Cody Geary, right image courtesy of Shinnosuke Seki.*

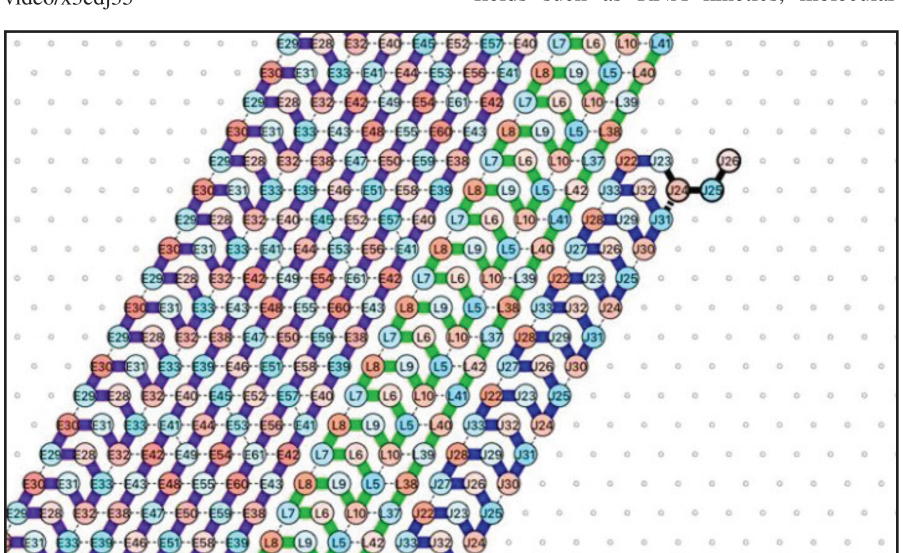


Figure 3. *Glider, a delay-3 oritatami system folding a rigid directional structure motif. View the complete animation at http://www.dailymotion.com/video/x3cdj35.*

References

[1] Beyer, A.L., Bouton, A.H., & Miller Jr., O.L. (1981). Correlation of hnRNP structure and nascent transcript cleavage. *Cell, 26*(2), 155-165.

[2] Cook, M. (2004). Universality in elementary cellular automata. *Complex Systems, 15*, 1-40.

[3] Feynman, R.P. (1996). *Feynman Lectures on Computation*. Cambridge, MA: Perseus Books Group.

[4] Geary, C., Meunier, P.-É., Schabanel, N., & Seki, S. (2016). Programming biomolecules that fold greedily during transcription. In *Proc. 41st International Symposium on Mathematical Foundations of Computer Science* (*MFCS 2016*). Vol. 58 of *Leibniz International Proceedings in Informatics* (pp. 43:1-43:14).

[5] Geary, C., Rothemund, P.W.K., & Andersen, E.S. (2014). A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science, 345*, 799-804.

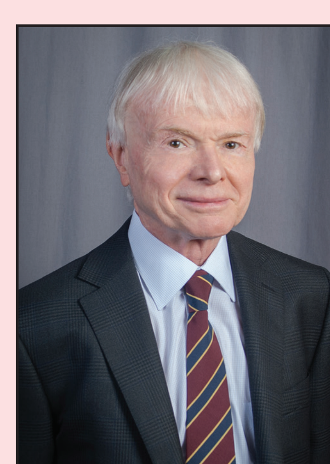
[6] Merkhofer, E.C., Hu, P., & Johnson, T.L. (2014). Introduction to cotranscriptional RNA splicing. In K.J. Hertel (Ed.), *Spliceosomal Pre-mRNA Splicing: Methods and Protocols* (pp. 83-96). Vol. 1126 of *Methods in Molecular Biology*, New York, NY: Springer.

[7] Wu, Z., Murphy, C., Callan, H.G., & Gall, J.G. (1991). Small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins in the amphibian germinal vesicle: Loops, spheres, and snurposomes. *Journal of Cell Biology, 113*(3), 465-483.

[8] Xayaphoummine, A., Viasnoff, V., Harlepp, S., & Isambert, H. (2007). Encoding folding paths of RNA switches. *Nucleic Acids Res., 35*(2), 614-622.

Shinnosuke Seki is an assistant professor in the Department of Computer and Network Engineering at the University of Electro-Communications in Tokyo, Japan. His research interests are in theoretical and computational models of molecular self-assembly.

Thomas J.R. Hughes Receives the SIAM/ACM Prize in CSE



Thomas J.R. Hughes of the University of Texas at Austin.

Thomas J.R. Hughes was awarded the 2017 SIAM/ACM Prize in Computational Science and Engineering on March 2 at the 2017 SIAM Conference on Computational Science and Engineering (CSE17) in Atlanta, Ga.

Hughes is the Peter O'Donnell Jr. Chair in Computational and Applied Mathematics and a professor of aerospace engineering and engineering mechanics at the Institute for Computational Engineering and Sciences at the University of Texas at Austin.

The prize honors Hughes for his pioneering work on finite element methods for partial differential equations. His work is used worldwide in engineering design and simulation, and has impacted every field of science that uses finite element methods. Hughes has also made pioneering contributions to the seamless integration of modeling methodologies with design representations. He has created entirely new fields of research, including stabilized methods, variational multiscale methods, and isogeometric analysis, and continues to lead their development.

"Almost my entire research career has been devoted to the development of computational methods used in engineering and science, so the SIAM/ACM Prize in Computational Science and Engineering, which is the major distinction in the field, represents in many ways the culmination of my life's work," Hughes said. "It means a great deal to me to be selected by my peers from among numerous outstanding candidates. I am truly honored and humbled to receive this award."

Climate Prediction, Exascale Computing, and Time Parallelism

By Paul Davis

New computing architectures promise to deliver more accurate weather predictions faster than ever before. But achieving dramatic reductions in wall-clock time using these new machines demands "disruptive algorithms," according to Beth Wingate of the University of Exeter, who gave an invited talk at the 2017 SIAM Conference on Computational Science and Engineering (CSE17), held in Atlanta, Ga., this February.

Wingate and her colleagues are among the vanguard of computational scientists developing parallel-in-time algorithms. These algorithms augment the classic paradigm of spatially parallel computing—assigning each of a multitude of processors to compute simultaneously the governing model equations at one or more distinct nodes of the spatial grid—with the intuitively disruptive idea of simultaneous computation at multiple points in time. How indeed could time evolve in any manner but serially, one moment after another?

Suppose each point of a temporal grid were the starting time for a distinct initial-value problem with its own dedicated processor(s). One could solve that ensemble of problems simultaneously—parallel in time—matching the solutions sequentially at the temporal grid points by adjusting the initial value of the outgoing solution to match the final value of its incoming predecessor.

Although there could be some fearful devils hiding in the details of such matching (more about them later), real run times might be cut dramatically. A long sequence of tiny time steps could be folded over on itself many times, potentially reducing the wall-clock time required for the computation. Of course, the overhead of matching at the temporal folds and other particulars could eat up those real-time savings if they were handled carelessly.

But thoughtfully constructed algorithms that are parallel in time do have the power to cope with the multiple time scales inherent in models for climate and weather prediction; indeed, multiple time scales are a challenge common to many large-scale computational models. Phenomena that oscillate rapidly, such as ocean vortices, demand small time steps to accurately track their motion. Dissipative phenomena, like cooling, may need a small time step to track an initial rapid decline in temperature but a much larger step as the rate of cooling decreases.

Although these disparate time scales may only present themselves at a few locations within the geographic region being modeled, the constraint of the smallest time step can impose itself much more widely. In a further twist of the tiny-time-step dagger, finer spatial resolutions necessitate even smaller time steps in order to maintain numerical stability.

The practical outcome is discouraging. As Wingate observed, despite the promise in the newest machines of vastly increased numbers of processors with more concurrency—the potential to do millions of calculations at the same time—total clock time needed to compute current weather and climate models may not decline significantly; finer spatial grids and better resolution of oscillatory phenomena will force more smaller time steps to model the same period of time. While the results may offer better spatial and temporal accuracy, they might not reach an investigator's desk any sooner — or demand less energy to compute.

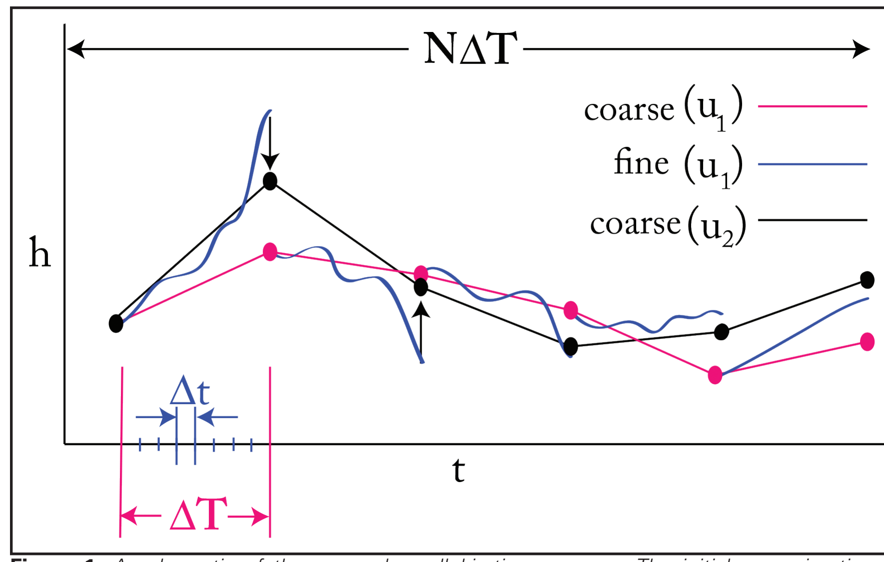


Figure 1. *A schematic of the parareal parallel-in-time process. The initial approximation to the solution of an initial-value problem on the coarse ΔT time grid is traced in pink. The subsequent corrections computed in parallel using the fine steps Δt are in blue. The improved coarse grid approximation is in black. Image adapted from [2].*

Mathematical and computational ingenuity are essential to realizing the promise that parallel-in-time algorithms will reduce overall execution time by simultaneously taking many small time steps.

In 1964, Jürg Nievergelt suggested the first explicitly parallel approach to solving an initial-value problem for an ordinary differential equation [1], though he was obviously unaware of the challenges and opportunities posed by today's most advanced architectures. He divided the required time interval into a series of coarse subintervals to permit a parallel, then a serial step: first compute in parallel a family of solutions for different initial values at the beginning of each subinterval, then sequentially interpolate solutions from one subinterval to the next to construct the solution over the original time interval.

The short time steps necessary for either numerical stability or accurate resolution of oscillations are still required to solve the coarse grid problems. Solving them simultaneously—in parallel—can save significant time if sequential matching at the grid points is not too costly.

The modern parallel-in-time paradigms developed by Wingate and her collaborators are rooted in a method called *parareal,* introduced in 2001 by Jacques-Louis Lions, Yvon Maday, and Gabriel Turinici. In simplest form, the parareal method also coarsely subdivides the original time interval, but it begins with an approximate solution computed on the coarse grid using the backward Euler method. These backward Euler values are the initial values for solution in parallel of each coarse grid initial-value problem. The backward Euler method propagates the discrepancy between the initial and final values at each subinterval interface. The coarse grid initial values are corrected accordingly, and these initial-value problems are solved again in parallel [1], as illustrated schematically in Figure 1. Ultimately, the iterative process of implicit solution and updating of initial values on the coarse grid can be reformulated as an incomplete Newton's method.

These parallel-in-time approaches are not just intuitively clever. They are provably accurate; e.g., each repetition of the propagate-and-solve cycle of the parareal method increases its order of accuracy by one. A firm analytic foundation for these parareal-type methods is key to the approach that Wingate and colleagues such as Terry Haut of Lawrence Livermore National Laboratory [2] use to reduce overall run time for a set of one-dimensional shallow water equations, a standard test problem for potential weather and climate algorithms.

Asymptotically, as the ratio of slow to fast frequencies approaches zero, the fast frequencies can be swept out of the problem. Wingate represents that process with a time average of a version of the underlying nonlinear operator; she emphasizes that the operator, not the solution, is averaged. Roughly speaking, her team's version of parareal uses the time-averaged operator/asymptotic approximation to provide the coarse time grid solution and an exponential integrator for the fine grid solution that corrects the coarse grid values. This algorithm can use coarse time steps that are as much as 50 times greater than the explicit time step limitation — and 10 times larger than the next best parallel method. More importantly, this approach incorporates the dynamic variations in time scales that are common in climate models; the ratio of slow to fast frequencies is not always small. A moving, variable-width time average window accommodates these situations.

Wingate's student Adam Peddle has both analytic and computational evidence (presented in a minisymposium[1] at CSE17) of a "Goldilocks sweet spot" in the number of parareal iterations versus the ratio of the width of the time-averaging window to the coarse grid spacing. When frequency ratios are not small, Peddle's results show an optimum number of parareal iterations, a point where "the averaging mitigates the oscillatory stiffness."

Wingate envisions an evolutionary integration of time parallelism into new machine architectures. "The immediate need is to get current models running on the new machines," she said. "Time parallelism will be in the background." Then new models, relatively simple at first, will become available to scientists studying simple, more fundamental problems. Today's young researchers will carry forward those models with novel algorithms on new machines, thereby advancing the methods that underpin our most important applications.

Wingate's CSE17 presentation is available from SIAM either as slides with synchronized audio, or as a PDF of slides only.[2]

Acknowledgments: *SIAM News* would like to acknowledge Beth Wingate for her review of this article.

References

[1] Gander, M.J. (2015). 50 years of time parallel time integration. In T. Carraro, M. Geiger, S. Korkel, & R. Rannacher (Eds.), *Multiple Shooting and Time Domain Decomposition*. Springer International Publishing.

[2] Haut, T., & Wingate, B. (2014). An Asymptotic Parallel-in-Time Method for Highly Oscillatory PDEs. *SIAM J. Sci. Comput., 36*(2), A693-A713.

Paul Davis is professor emeritus of mathematical sciences at Worcester Polytechnic Institute.

[1] http://meetings.siam.org/sess/dsp_talk.cfm?p=81249

[2] https://www.pathlms.com/siam/courses/4150/sections/5821

Adaptive Mesh Refinement: An Essential Ingredient in Computational Science

By Paul Davis

Adaptive mesh refinement may be to computational science and engineering (CSE) what Bolognese sauce is to Italian cooking: part of many meals and integral to the repertoire of most cooks, nearly all of whom are happy to share and serve their special recipes. In that spirit, adaptive mesh refinement was widely served at invited presentations, minisymposia, and poster sessions at the 2017 SIAM Conference on Computational Science and Engineering (CSE17) in Atlanta, Ga., this February; "adaptive mesh refinement" and "AMR" are mentioned on 52 distinct occasions in the meeting's abstracts!

Like Bolognese, AMR has but a few simple ingredients. The trick is managing them well — across many, many pots on a very large stove. AMR aims to efficiently accommodate the vast variations in scale inherent in most physical phenomena by using smaller scales only when and where required.

As the components of a set of partial differential equations' approximate solution advance in time, AMR refines selected portions of the spatial grid to meet predetermined error bounds (and perhaps takes smaller time steps, too.) The mesh refinement is dictated adaptively by a local estimate of the error in the computed solution; the sous-chef de précision is automated — embedded in the code. Clever bookkeeping tracks the mesh hierarchy and the approximate solution values as mesh points appear and disappear.

Abel, Adaptive Methods, and Astrophysics

Tom Abel of Stanford University began his invited presentation, "Making Sense of the Universe with Supercomputers," with a cosmological problem that incorporates the "mother of all scales:" tracing the formation "via ab initio modeling" of the very first stars from the soup of hydrogen, helium, dark matter, etc. that constituted the universe a mere 400,000 years after the Big Bang.

"We can't experiment," Abel remarked dryly, because "there are no black holes in the laboratory." But, he suggested, we do know what that soup looked like 400,000 years post-Big Bang because we can measure now the background radiation it emitted then. "Think initial conditions," Abel said. Of course, the point of the computational modeling is determining what transpired in the "dark ages" between the time of that discordant brew and the nearby universe we observe today.

The scales in this computation are certainly daunting. Spatial variations on the order of 10^{12} and a dynamic range up to 10^{15} require AMR in both space and time. The model itself is another soup, with multi-physics ingredients such as gravity, gas dynamics, gas chemistry, radiation, and magneto-hydrodynamics.

Abel and his colleagues' computational package of choice is Enzo, "a community-developed AMR simulation code designed for rich, multi-physics hydrodynamic astrophysical calculations."[1] With 300,000 lines of C++ and Fortan77, the code itself is a massive undertaking of computational engineering, never mind the computational science underlying core algorithms (like AMR) that it implements, or the astrophysics of the daunting questions it seeks to answer.

Figure 1 (on page 1) illustrates one of the successes of these efforts: isodensity and isothermal profiles computed at three levels of spatial resolution during the evolution of the very first stars. Abel noted that while average properties like mass and temperature converge reasonably well in these computations, measures of vorticity, turbulence, and magnetic field generation are less accurate.

Moving from the birth of the first stars to the evolution of the second generation adds the computational complexity of ray-tracing. Further up the scientific mountain being scaled by Abel and his colleagues—and even more spectacular—are predictions of the distribution of dark matter in the universe. These suggest that "galaxies are arranged in a cosmic web of voids, sheets, filaments, and holes," none of which can be predicted analytically. Public television programs often use visualizations of aspects of the universe's evolution contributed by Abel and his colleagues, some of which he included in his talk.

"Few problems of astrophysics can be addressed in the lab," Abel said, emphasizing the field's reliance on computation. Since complex solutions require more memory, efficient adaptive tools like AMR are essential, although AMR is but one tool among many that his teams employ. In its development of Enzo, the astrophysics community was also an early adopter of many of the ground rules of distributed code development—like version control—which are now commonplace.

[1] www.enzo-project.org

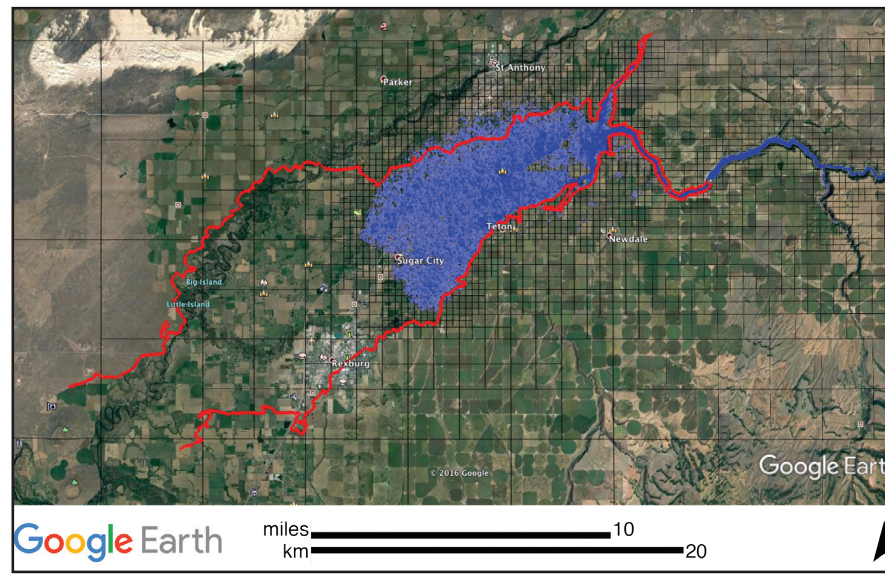


Figure 2. *Simulation of flooding resulting from the 1976 Teton Dam failure in eastern Idaho, done using ForestClaw, a parallel adaptive quadtree code for patch-based adaptive mesh refinement. The flood waters reached Sugar City (shown on figure) with a 15 foot wall of water inundating the city. The thicker red line is a digitized flood boundary, taken from historical records, and shows good agreement between the ForestClaw results and the historical record. The simulated flood arrival times (not shown) at Sugar City, Rexburg, and beyond also show excellent agreement with historical data. © Google, Digital Globe. Image credit: Donna Calhoun.*

Almgren and the Next Generation of AMR

Ann Almgren of Lawrence Berkeley National Laboratory focused her invited presentation exclusively on the development of the next generation of AMR algorithms. Her core message, aimed at a CSE17 audience of AMR cognoscenti, is that block-structured AMR provides natural opportunities for parallelism because it manages data efficiently.

In principle, AMR could arbitrarily subdivide parts of individual coarse grid cells higgledy-piggledy—one here, another there—so that the refined grid presents no discernable regular pattern within the original coarse grid. Block-structured (or tile-based) AMR regards the individual grid cells as a set of non-overlapping tiles of fixed size. An individual tile is refined by subdividing it uniformly and by appending so-called ghost cells to store solution values from neighboring tiles that the algorithm will need to advance the refined tile's solution in time.

Almgren argued that block-structured AMR offers "a natural framework for reducing memory use," and that its "infrastructure naturally supports hierarchical parallelism." Put crudely, if the code can track the grid refinements of AMR and the coders know enough about the machine and the model being solved, then code and coders can together make efficient decisions about using storage and assigning tasks to nodes of the machine. Of course, the code's "decisions" are adaptive algorithms, which coders decide to incorporate during design and development.

Almgren comprehensively addressed six aspects facing next-generation AMR: single-core versus single-node performance, programming models, load balancing, synchronicity, the possibility of new equations and corresponding new algorithms, and the trade-offs between in situ and in transit visualization and analysis.

See **Adaptive Mesh Refinement** *on page 8*

A Model for an Applied Mathematics Internship Program

By Nadia Benakli and Jonathan Natov

The Department of Mathematics at the City University of New York's New York City College of Technology (City Tech) began offering an applied mathematics major in 2004. The program has since found much success, with enrollment growing from only seven majors at its onset to approximately 100 in 2016. Our internship requirement is a key part of this success.

City Tech's students come from more than 100 different countries. Many are first-generation college-goers, and often need employment immediately after graduation. While some of the internships are paid, it is more essential that the work experience helps students reach their career objectives. Six of the 12 interns in our spring 2016 cohort remained at the institution where they completed their internships, and 10 of the 12 intern supervisors indicated that their agency, or a similar organization, would likely hire the intern. This suggests that even more interns could have found employment at their internship sites, but decided to go on to graduate school or look for another position.

CAREERS IN MATHEMATICAL SCIENCES

In the following sections, we discuss our management and assessment of the internship program. We also provide some case studies, which highlight the utility of mathematics in atypical settings.

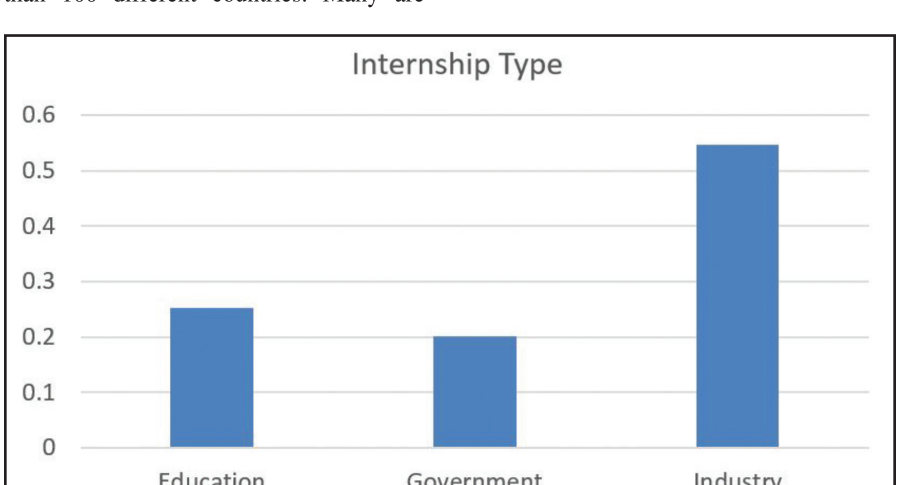


Figure 1. *Breakdown of internship types for students in the Department of Mathematics at City Tech. The data is based on 139 internships over the past 12 years. Figure credit: Nadia Benakli and Jonathan Natov.*

Finding Internships

Students in City Tech's applied math program are required to complete two internships. Students submit statements of career objectives at the beginning of the process, which help determine appropriate internship opportunities. They are responsible for finding appropriate positions, with some help from faculty, and can also access City Tech's Professional Development Center for possible opportunities and on-campus recruiting events.

While some students complete their internships in "conventional" host organizations, such as financial institutions, educational foundations, or research labs, many internship opportunities come as a surprise, as evidenced in the following case studies. Figure 1 presents a breakdown of students' internship types, based on 139 internships over the last 12 years.

Case Studies

Many of our interns have discovered interesting opportunities for mathematics and consequently broadened our perspective of what an internship might look like. Their successes reinforce our belief that mathematics has a role to play in almost every industry.

Transportation. One applied mathematics major obtained a paid job working on signal repairs for the New York City Metropolitan Transportation Authority (MTA). Many of our students work part time or full time while pursuing their degrees, and City Tech has an established relationship with the MTA.

The opportunity arose during the performance of a mandated test to determine the condition of batteries, which identify the position of trains in the event of a power outage. By collecting data and using a simple regression analysis, the student estimated the longevity of battery life. This work became the basis for his first internship.

Given the vast amounts of data that have been collected but not yet analyzed, the intern believes that his data analysis skills will yield many opportunities to improve the system.

Transportation: Railroad. Another transportation-related example involves a

See **Internship Program** *on page 7*

Internship Program

Continued from page 6

student who worked at a leading railroad company to support his family. Though his job focused on routine tasks, he astutely noticed that a certain switch often broke down during system failures, causing significant problems and passenger delays. He downloaded a student version of a statistical application, taught himself the necessary commands, collected data, and presented the analysis to his supervisor.

By replacing an inexpensive switch, the railroad significantly reduced the system failures. In response, the railroad management hired the student as a systems analyst and supported him while he completed a graduate program.

Call Center. Yet another student wished to ultimately find employment with a financial institution. Her internship experience began while working for a city agency, answering calls and directing callers to the appropriate department. While this did not qualify as an internship, she took the initiative to keep track of the approximate number of calls the center received. She then created a statistical model to determine a confidence interval that estimated the likely number of callers within an hour. Her work showcased inferential statistics and impressed her supervisor.

This intern then moved on to a position as a budget analyst for a car rental agency, which led to a job as an analyst at a prestigious financial institution.

Pharmacy. When one of our majors began working for a pharmaceutical company, his job essentially involved running errands. However, he looked for opportunities to use his skills. He noticed that the company tended to purchase relatively small amounts of chemicals, and wondered if purchasing in bulk could save money. But price fluctuations cause uncertainty, and the company was unfamiliar with the cost of larger quantities of chemicals.

The student produced a price estimate using a regression analysis, which closely matched the actual price the company would have to pay. His applied mathematics skills promoted him from running errands to running budget analyses.

Teachers' Union. While a teachers' union might not seem like an obvious place for an applied math student, one of our interns found unexpected success. The union in question is part of an umbrella organization to which it must pay dues. At the start of a new billing cycle, the union can choose from a few different payment formulas.

Formulas are based on the number of full- and part-time faculty members, higher education officers, and their corresponding ranks. Computing the cost involved a large data set (in the tens of thousands); unpredictable fluctuation of the numbers was an added complication. Our intern helped create a pricing model to accurately determine the cost of each payment option. In so doing, he helped the union save hundreds of thousands of dollars each year.

Data Science. A student found a position at a company that analyzes television viewership data. The company developed both a proprietary application to quantify qualitative data and a benchmarking application to test the proprietary application's success.

After graduation, the student landed a full-time job with the company, which subsequently took on two more interns. As a result of our students' many successes, companies are beginning to approach us for interns. With time, we expect growing demand for our applied math majors.

Assessing the Internships

We assess internships based on a supervisor's evaluation, a log of the intern's activities, a final written report, an oral presentation, the student's professionalism, and the significance and relevance of their work to the overall goals of the organization.

Interns are expected to behave as professionals, which involves using terminology appropriate for industry, writing professional documents, respecting deadlines, keeping scheduled appointments, and conducting themselves with integrity and respect.

Written reports submitted by interns include information on the host organization; an overview of the intern's responsibilities, duties, and overall experience; and insights as to how the intern's contributions benefited the company. Keeping a work log helps interns evaluate their work and monitor their progress while recording their weekly tasks and new skills.

Towards the end of the semester, interns present their work. They focus on essentials, respond to audience questions, and are expected to use their limited time wisely.

Finally, we also consider the significance of internship work for overall evaluation. By the end of their second internship, students should be prepared to meet their career objectives.

Ensuring a Strong Curriculum

The internship experience is the heart of City Tech's applied mathematics program, and serves three essential functions. First, it helps graduates find meaningful employment. A successful internship provides strong evidence of future success in industry, which assists our graduates as they compete against candidates from more prestigious institutions.

Secondly, it ensures that our curriculum meets industry's current demands and incorporates current applications of mathematics that are of interest and relevant to industry.

Ultimately, internships provide a way to assess the strengths and weaknesses of our program. For example, supervisors have generally been impressed with our interns' mathematical skills. However, ratings of their communication skills were typically lower. In response, we modified our math modeling courses to include more written reports and oral presentations. This change has correlated with improved ratings.

We strongly recommend that applied math programs require internships, in order to foster strong mathematical training and keep students competitive. Our experience has shown that internships are manageable and crucial for a career-oriented program.

Further Reading

[1] ACT, Inc. (2000). *Workplace Essential Skills: Resources Related to the SCANS Competencies and Foundation Skills*. Presented to the U.S. Department of Labor, Employment and Training Administration & the U.S. Department of Education National Center for Education Statistics. Retrieved from https://wdr.doleta.gov/opr/fulltext/00-wes.pdf.

[2] Fregosi, D. (2012). Should My Salary Be Higher After Graduation If I Have An Internship? *EzineArticles.com*. Retrieved from http://EzineArticles.com/7074831.

[3] Korkki, P. (2011, March 25). The Internship as Inside Track. *The New York Times*. Retrieved from http://www.nytimes.com/2011/03/27/jobs/27searches.html?_r=0.

[4] Long Island Press. (2010). Why Internships Are Essential. *Long Island Press*. Retrieved from http://archive.longislandpress.com/2010/01/07/why-internships-are-essential/.

[5] National Association of Colleges and Employers. (2013). 2013 Internship and Co-op Survey. Retrieved from https://www.naceweb.org/uploadedfiles/content/static-assets/downloads/executive-summary/2013-internship-co-op-survey-executive-summary.pdf.

[6] Sterrett, A. (Ed.). (1996). *101 Careers in Mathematics*. Washington, D.C.: Mathematical Association of America.

Nadia Benakli is an associate professor of mathematics at New York City College of Technology of the City University of New York (CUNY). She is the coordinator of the applied mathematics internship program. Jonathan Natov is a professor of mathematics at NYC College of Technology. He coordinated the applied mathematics program from 2004 to 2014.

Broader Engagement Program Returns to CSE17 with a Focus on Community Engagement

By Debbie McCoy and Mary Ann Leung

The Sustainable Horizons Institute's Broader Engagement (BE) program[1] aims to provide a rich scientific agenda, mentoring, and career and professional development opportunities to students, faculty, and professionals aspiring to widen their experience in research-based professional activities. Computational science and engineering (CSE) is at the forefront of investigation and engineering design in areas ranging from aerospace and automotive industries to biological, chemical, and semiconductor technologies. CSE's prevalence in these fields and reliance on advanced modeling and simulation make it the ideal backdrop for projects such as BE.

The program therefore returned to the 2017 SIAM Conference on Computational Science and Engineering (CSE17) in Atlanta, Ga., this February, having made its debut at the 2015 CSE conference.[2] This venture brings a diverse group of students, faculty, and professionals to the conference, thus involving them in conference activities and increasing the CSE community's knowledge and support of diversity and inclusion.

"The unique aspect offered by the BE program is that it allows individual researchers or academics to contribute to the goals of diversity at the grassroots level," said Hans De Sterck, professor of computational and applied mathematics at Monash University and chair of the SIAM Activity Group on Computational Science and Engineering. "I volunteered because I felt it is important to try to directly and personally contribute to making SIAM events like CSE17 more accessible to early career and underrepresented groups. Increasing diversity benefits both individuals and the profession."

At CSE17, the BE program focused on engaging the computational science and engineering professional communities with students and early career scientists. "I always wanted to help and engage with people, but BE finally provides an actual mechanism to do this, and that's great!" De Sterck said. "I was also curious about how BE activities are set up and how they work, just to know what kinds of things can be done and perhaps used at other events; I believe this kind of initiative can be valuable in many settings."

During CSE17 registration, over 270 members of the CSE community indicated an interest in learning about BE volunteer opportunities. More than 50 CSE community members helped out as mentors, leaders for guided affinity groups, and volunteers for the Student Opportunities Lab (informal roundtable discussions about career/professional development topics). In addition, Google employees served as ambassadors for BE participants and accompanied them on a site visit to Google's Atlanta-based Fiber Academy.

Broader Engagement (BE) participants toured Google's Atlanta-based Fiber Academy during the 2017 SIAM Conference on Computational Science and Engineering. Photo credit: Mary Ann Leung.

"I volunteered to lead a guided affinity group because I appreciate the opportunities I've had to attend beginners' forums and tutorials," said Stephen Wood, postdoctoral researcher in computational engineering at the University of Tennessee's (UT) Innovative Computing Laboratory and the UT-Oak Ridge National Laboratory Joint Institute for Computational Sciences. Wood led a guided affinity group on uncertainty quantification (UQ) for computational fluid dynamics in sustainable energy applications. "I saw the opportunity as a moment to pay forward the sage guidance I've received from more experienced practitioners," he said. "It was a pleasure to interact with inquisitive students from a variety of disciplines. The best part of the experience was listening to students discuss how they could apply UQ to their work; the openness of their questions and helpful responses to each other continue to inspire me."

While 33 BE participants were sponsored by the Sustainable Horizons Institute, 23 others paid their own way to be part of the program. Sponsored participants—11 undergraduates, 16 graduate students, and six early career professionals—were representative of 30 institutions, 10 of which are historically black colleges and universities and minority-serving institutions.

"I would not be where I am today without the amazing mentors who have encouraged, advised, and inspired me along the way," said Julianne Chung, assistant professor in the Department of Mathematics and the Computational Modeling and Data Analytics Division at Virginia Tech. "Broader engagement programs provide opportunities for me to give back to the community."

The BE program at CSE17 matched students and early career participants with mentors to create and foster networks between students, early career scientists, established researchers, and leaders in scientific communities. "BE's mentorship program, where mentors are paired with BE protégés based on background, interests, and needs, is an excellent way to develop and encourage a strong and diverse workforce of computational scientists," Chung said. "Navigating

See **Broader Engagement** *on page 11*

[1] http://shinstitute.org/siam_cse17_be/
[2] https://sinews.siam.org/Details-Page/life-is-a-tensor-pilot-program-aims-at-expanding-siam-impact

Participants of the Broader Engagement (BE) program at the 2017 SIAM Conference on Computational Science and Engineering, held in Atlanta, Ga., this February. Photo credit: Nitin Sukhija.

Adaptive Mesh Refinement

Continued from page 6

Load balancing, for example, is usually based on the number of cells. But if the particles in the simulation are unevenly distributed on the grid, then redistributing work by particle might be more efficient, depending on the relative costs of computation and data transfer.

Although Almgren acknowledges that "'Synchronicity' means different things to different people," she uses it to label a range of coordination trade-offs that might be productive. Some could take place at a very low level, entirely invisible to the application, e.g., floating point computations on some cells while doing simple bookkeeping on others. Other coordination strategies might be quite brazen—say, changing the order of high-level tasks within the algorithm—provided one tile doesn't step far ahead of others in time, a situation that would postpone solution updates and inflate memory demands.

New algorithms might be implemented differently to remove synchronization points like norm calculations. Or new algorithms might solve different equations, perhaps permitting asynchronous combustion or a particle description of a fluid embedded within a continuum model. In the face of so many possibilities, Almgren encouraged computational scientists to "Keep the options open!"

AMR Elsewhere

AMR appeared throughout CSE17 in a myriad of sessions. As but one example, it was often part of simulations of surface flooding using AMR-based tools from the Conservation Laws Package (Clawpack), a community-supported collection originally developed by Randall LeVeque of the University of Washington. LeVeque organized a six-poster session describing some of the Claw packages' applications at CSE17.

Donna Calhoun[2] of Boise State University; Yu-Hsuan Melody Shih, then at Columbia University, now at Boise State; Kyle Mandli of Columbia; Carsten Burstedde of the University of Bonn; and collaborators from Idaho National Laboratory used the GeoClaw[3] extension of ForestClaw,[4] a parallel adaptive quadtree code for patch-based AMR, to simulate the June 1976 flooding that ensued from the catastrophic failure of Idaho's Teton Dam. Eleven people were killed and approximately 2 billion dollars of property destroyed as the flood spread 88 kilometers downstream in the eight hours following the dam's collapse.

One of the investigators' goals was assessing the suitability of these tools "for a potential study of flooding of nuclear power plants." In their simulation, they sought to match records of the crest's arrival times and the flood's geographic spread, though not the dynamics of the dam's actual failure.

Calhoun, Shih, and Mandli found that they needed a better model of the dam's burst to accurately modulate the water's initial flow and volume. Changes in these initial conditions affected the agreement between simulated and actual arrival times and between the predicted and actual spread of flood waters (see Figure 2, on page 6). Despite the uncertain initial data, the algorithm performed well; the team could see the grid refinement as the water advanced. About 50-70% of CPU time was spent actually solving the problem.

In a related study, Mandli, working with Colton Conroy and Jiao Li of Columbia, used patch-based AMR to solve the shallow water equations with added terms in order to assess the risk of storm-surge flooding in coastal cities. The trio sought efficient, reliable answers to some of the fundamental questions of flood management design and forecasting: How high should we build surge protection barriers? Will the water deflected by the barriers cause flooding elsewhere? They found partial answers by solving a Riemann problem over a region where barriers were modeled as thin walls at cell boundaries, although much work remains.

A Large Cast

Of course, AMR is but one among many arms that CSE has built for itself and repeatedly flexed in the service of other areas of human endeavor, ranging from the astronomic to the atomic. AMR's frequent appearances in so many settings at CSE17 make it a compelling lead character—but only one among a very large cast—in this story about computational science and engineering as a fruitful and rapidly-growing part of applied mathematics.

Abel's[5] and Almgren's[6] CSE17 presentations are available from SIAM either as slides with synchronized audio, or as PDFs of slides only.

Paul Davis is professor emeritus of mathematical sciences at Worcester Polytechnic Institute.

[2] Calhoun maintains an extensive list of adaptive mesh resources at http://math.boisestate.edu/~calhoun/www_personal/research/amr_software/
[3] www.geoclaw.org
[4] www.forestclaw.org

[5] https://www.pathlms.com/siam/courses/4150/sections/5827
[6] https://www.pathlms.com/siam/courses/4150/sections/5833

SIAM is pleased to announce the newly selected Class of SIAM Fellows—a group of distinguished members of SIAM who were nominated by their peers for exceptional contributions to the fields of applied mathematics and computational science. Please join us in congratulating these 28 members of our community.

siam Fellow

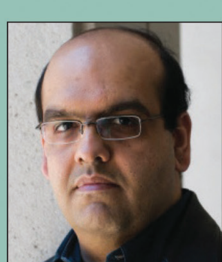
Announcing the 2017 Class of SIAM Fellows

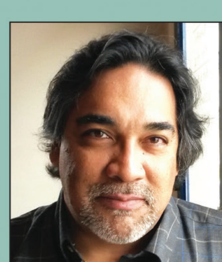

Zhaojun Bai
University of California, Davis


Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems


Angelika Bunse-Gerstner
Universität Bremen

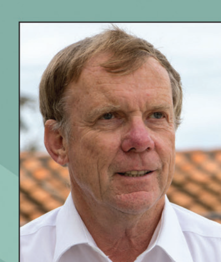

Emmanuel Candès
Stanford University


Rama Cont
Imperial College London


Ricardo Cortez
Tulane University


Lieven De Lathauwer
KU Leuven


Bart De Moor
KU Leuven


Andreas Griewank
Yachay Tech University


Helge Holden
Norwegian University of Science and Technology


Panayotis Kevrekidis
University of Massachusetts, Amherst


Vipin Kumar
University of Minnesota


Karl Kunisch
Karl Franzens Universität Graz

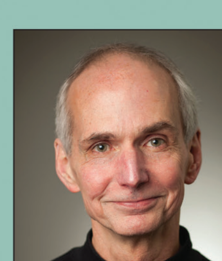

Monique Laurent
Centrum Wiskunde & Informatica

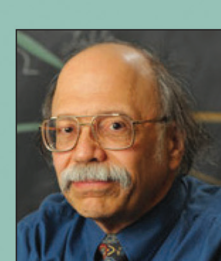

Mark A. Lewis
University of Alberta

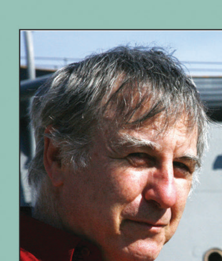

Lois Curfman McInnes
Argonne National Laboratory


Igor Mezić
University of California, Santa Barbara


Michael Kwok-Po Ng
Hong Kong Baptist University


James Renegar
Cornell University


Andrew J. Sommese
University of Notre Dame


Joel H. Spencer
Courant Institute of Mathematical Sciences, New York University


Gabor Stepan
Budapest University of Technology and Economics


Daniel B. Szyld
Temple University


Jean E. Taylor
New York University and Rutgers, The State University of New Jersey


Marc Teboulle
Tel Aviv University


J.A.C. Weideman
Stellenbosch University


Carol S. Woodward
Lawrence Livermore National Laboratory


Kevin Zumbrun
Indiana University

siam. SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS

www.siam.org/prizes/fellows

Nominate a colleague for the 2018 Class of Fellows. Nominations close November 6, 2017.

When Big Data Algorithms Discriminate

BOOK REVIEW
By James Case

Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *By Cathy O'Neil. Crown Publishing, New York, NY, September 2016. 272 pages, $26.00.*

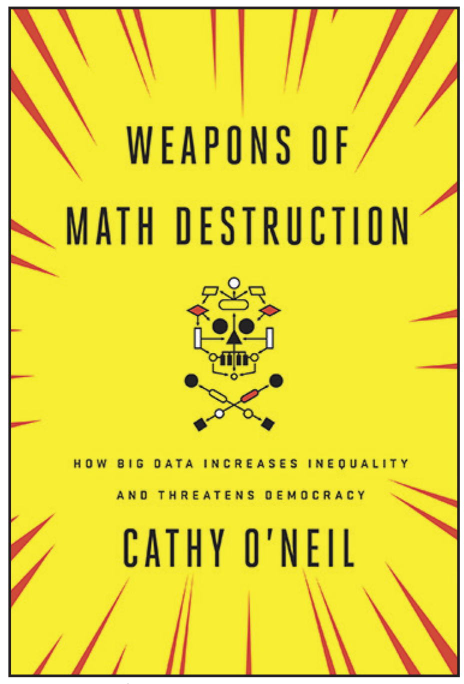
Cathy O'Neil discovered prime numbers on her own, at an early age. At 14 years old, she mastered the Rubik's Cube while attending summer math camp. Following undergraduate work at the University of California, Berkeley, O'Neil earned a Ph.D. from Harvard University in 1999, with a thesis in algebraic geometry. Next came a postdoctoral position at the Massachusetts Institute of Technology, followed by a tenure-track professorship at Barnard College/Columbia University. In 2007, seeking excitement, she elected to try her hand at finance with hedge fund D.E. Shaw.

After four of the most tumultuous years in Wall Street history, O'Neil became convinced that the computer programs used to scour the global economy for promising investment opportunities were partly to blame for the housing crisis, the collapse of major financial institutions, the rise of unemployment, and other societal plights. Moreover, she began to suspect that data-driven finance was but a small part of an emerging "big data economy," with limitless potential for good or ill.

A computer program can speed through thousands of résumés or loan applications in a second or two. Not only do the machines save time and money, they also treat everybody the same way — meaning they appear fair and objective in court.

Regrettably, programs written by humans nearly always encode at least a few of the biases, prejudices, and misconceptions harbored—consciously or unconsciously—by their creators. Moreover, the verdicts they render are all but impossible to appeal, since nobody really knows what makes the programs work the way they do. O'Neil maintains that, whether by accident or design, too many of these electronic decision-makers punish the poor and oppressed while further rewarding the already-rich. For ease of reference, she took to describing the more dangerous decision-making programs as "weapons of math destruction," or WMDs for short.

O'Neil's opening example involves a procedure meant to improve Washington D.C.'s school system. In 2007, the new mayor set out to reform the district's underperforming schools. Only eight percent of the system's eighth graders were performing at grade level in math, and barely half of those entering high school were soldiering on to graduation. Choosing to blame the teachers, the mayor decided that the solution was to identify and remove those that were incompetent. To that end, he created a powerful new post—chancellor of D.C. schools—and hired Michelle Rhee, a young but highly-regarded reformer, to fill it.



Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. By Cathy O'Neil. Courtesy of Crown Publishing.

Rhee engaged a Princeton, N.J.-based consulting firm called Mathematica Policy Research to construct a "value added" program known as IMPACT to measure each student's year-to-year learning progress. Teachers were then rated by their students' progress. At the end of the 2009-2010 school year, all teachers with IMPACT scores in the bottom two percent were fired. A year later, another five percent—206 teachers total—were terminated. Everything seemed to be going according to plan, including collateral damage.

Sarah Wysocki, a fifth grade teacher with two years of experience, had received nothing but positive feedback from her superiors and students' parents. One evaluation praised her attentiveness to the children, while another called her "One of the best teachers I've ever come in contact with." Yet her IMPACT score was dismal, obliging the district to fire her. How, she wondered, could this have happened?

Upon inquiry, Wysocki learned that her students' test papers from the previous year, which weighed heavily in their prior IMPACT scores, contained an unusual number of erasures. Had prior teachers changed answers to improve scores? Had they protected their own jobs, while costing Wysocki hers? Fortunately, Wysocki quickly landed another job in Virginia, where teachers are evaluated differently. She arrived with glowing letters of recommendation, while D.C. retained a possibly dishonest (and/or incompetent) teacher.

Wysocki's firing was an obvious miscarriage of justice, quickly recognized and soon corrected. But what of the other mishaps, the ones that went undetected? Wysocki could hardly have been the only qualified teacher to be fired. What, one wonders, do value-added assessments actually measure? Do they measure a teacher's ability to teach? Do they measure his or her impact on students? Or do they measure nothing at all? The Tim Clifford case suggests that the latter possibility is entirely too real.

Clifford was a middle school English teacher in New York City with 26 years of experience. When the city adopted a rating system similar to the one that cost Wysocki her job, he was shocked to learn that his initial rating was an appalling six out of 100. Clifford worried that with a few more such scores, even his tenured position might be in jeopardy. It also concerned him that poor scores for tenured teachers call into question the validity of the tenure system, already under fire from would-be reformers. So imagine his relief when, a year later—with no discernable change in his teaching methods—his rating rose to an enviable 96! How can one trust such a volatile performance index?

"In fact, misinterpreted statistics run through the history of teacher evaluation," O'Neil writes. She offers an imposing list of difficulties to overcome, and a litany of mistakes commonly made during the process.

Electronic decision-makers are, of course, by no means restricted to the education system. They are often used to decide which internet shoppers should see certain ads. Advertisers test different versions on (disjoint) samples drawn from a target demographic to learn which generate the greatest response. The winning version is shown to the entire audience of presumed "susceptibles" only after a number of these small-scale trials. Every purchaser of western garb, get-rich-quick schemes, weight-loss programs, exercise equipment, or exotic vacations is sure to be rewarded with "opportunities" to buy more such products.

Men and women in the armed services nearing completion of their tours of duty are routinely swamped with offers from for-profit universities, mainly because government loans are more easily obtained on their behalf. This, says O'Neil, is a particularly grievous use of electronic rating techniques because it encourages primarily poor, often poorly-educated, and easily-misled individuals to assume unrepayable quantities of debt in return for all-but-worthless credentials.

Many of O'Neil's complaints involve fairness and accuracy. Judges, she notes, often hand down more severe sentences to convicts deemed likely to become repeat offenders. They do so even though the likelihood in question is typically assessed electronically, and may be due to the offender's broken family or residence in a high-crime neighborhood. Although such information would be inadmissible in court due to its propensity to result in a verdict of guilt by association, it still counts against the defendant at sentencing time.

Similarly, loans are frequently denied to applicants considered liable to default, despite the fact that electronic credit rating algorithms often predicate their evaluations on applicants' residence in neighborhoods where jobs are likely to be temporary and/or loan defaults are unusually common. This also invites an unfair finding of guilt by association.

Finally, O'Neil laments the lack of concern regarding the accuracy of electronic rating schemes, despite the fact that their main purpose is to fairly narrow down the field of job seekers, loan applicants, or potential parolees at low cost. The fact that more labor-intensive evaluation schemes *might result in* slightly better workers, slightly fewer loan defaults, or slightly more law-abiding parolees counts for little beside the indisputable cost savings obtained with electronic screening.

O'Neil points out that this cavalier attitude toward accuracy is in marked contrast to the player evaluation schemes employed in professional sports, where even slightly better players can mean the difference between (profitable) winning and (unprofitable) losing records. Those electronic rating systems—like those that determine which ads will pop up on your computer screen—are continually monitored and improved. Too often, the ones applied to teachers, job seekers, loan applicants, and convicted criminals are not. Though easily available for the purchase of packages like IMPACT, funds for testing their accuracy are regrettably scarce.

O'Neil's important and fact-filled book will win no awards for suspense. One chapter tends to resemble another, since many of the same pitfalls await the use of electronic rating schemes in different fields of application. But by exposing the shortcomings of existing methods, *Weapons of Math Destruction* gives reason to hope that *demonstrably better systems* may yet be developed.

James Case writes from Baltimore, Maryland.

An Electrician's (or a Plumber's) Proof of Euler's Polyhedral Formula

MATHEMATICAL CURIOSITIES
By Mark Levi

Euler's famous polyhedral formula,

$$V - E + F = 2, \qquad (1)$$

describes the numbers of vertices, edges, and faces of a polyhedron "without holes," i.e., one that is sphere-like and in three dimensions. If the polyhedron has one hole, as in Figure 2, then we subtract 2 from the right-hand side of the formula, which becomes $V - E + F = 0$. The same thing happens for each additional hole (or, putting it differently, "handle on the sphere"). I will describe an argument based on electric circuits, leading to Euler's formula. I learned this beautiful idea from Peter Lax, and I therefore lay no claim to originality, except for any errors.

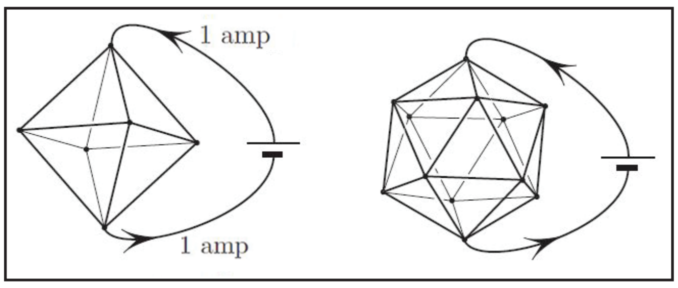


Figure 1. *V, E, and F are the numbers of vertices, edges, and faces of a polyhedron.*

Imagine our polyhedron as a wire frame, the edges being conducting wires, each of resistance 1 ohm, welded together at the vertices. Let us connect two vertices (chosen arbitrarily) to a battery, adjusting the voltage so as to drive the current of exactly 1 ampere. Now, nature will pick a specific value for each edge's current. In doing so, she obeys Kirchhoff's laws: the currents satisfy some equations that determine the currents. Let us take it for granted that

the number of unknown currents =
the number of independent equations.

This sentence is already Euler's formula in disguise! Indeed, the left-hand side is E, one unknown current per wire. For the right-hand side, Kirchhoff's laws state the following:

(i) The sum of currents entering each vertex is zero, giving V equations. But one of these equations is redundant, since it results from adding up all the others (I leave out the simple verification), yielding $V-1$ equations.

(ii) The sum of voltage drops around each face is zero. This gives F equations, one of which is the sum of the remaining ones and thus redundant, for the total of $F-1$ equations.

Summarizing,

$$E = (V-1) + (F-1),$$

which amounts to (1). For a polyhedron with a hole, as illustrated in Figure 2, we must add two more equations, expressing the fact that the voltage drop over each of two non-contractible circuits is zero, resulting in $E = (V-1) + (F-1) + 2$, or

$$V - E + F = 0.$$

Admittedly the proof is not rigorous as given, since, for instance, I did not eliminate the possibility of more redundant equations.

Although this proof would have sounded strange in Euler's pre-electricity days, it could be reformulated in plumber's terms by treating the polyhedron as a network of tubes (with porous blockages playing the role of resistors), currents as the mass flow per second, and voltages as pressures.

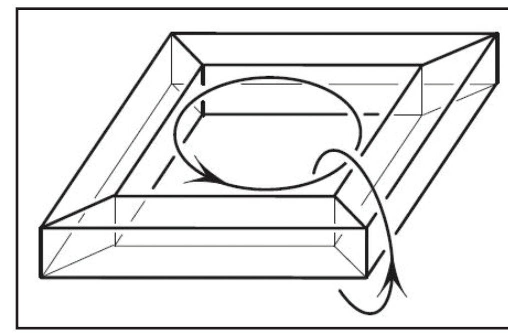


Figure 2. *A polyhedron with a hole; two non-contractible circuits are indicated.*

The figures in this article were provided by the author.

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

Broader Engagement

Continued from page 8

the ranks of academia or figuring out how to juggle family priorities with graduate studies can leave many feeling isolated. The mentorship program provides ample opportunities for discussions that may continue after the conference, where often times it is not just the mentees but also the mentors who walk away thinking, 'Wow, I'm not the only one who feels this way!'"

Participants presented posters and gave oral presentations on their research during the conference. The BE program won an award for best minisymposterium.

"I have gotten more out of this week than I have from all of the other conferences I have attended *combined*," a student participant wrote of the BE program. "It helps fill the gaps in my career preparation that one cannot get from usual graduate program activities. I especially appreciated the opportunity to be paired with a mentor, who helped me polish my résumé. It was fantastic!"

"The BE program as a microcosm of diversity in a larger, more homogeneous group was fantastic in that it provided a community for those of us who would have otherwise not been able to participate in the conference," another student shared. "It's a wonderful opportunity to combat impostor syndrome in those of us who tend to question whether we belong in academia/research careers."

Debbie McCoy is the director of programs at Sustainable Horizons Institute. Mary Ann Leung is president of the Sustainable Horizons Institute.

Charles L. Epstein

Leslie Greengard

Professional Opportunities and Announcements

Send copy for classified advertisements and announcements to: ***marketing@siam.org****; For rates, deadlines, and ad specifications visit* ***www.siam.org/advertising****.*

*Students (and others) in search of information about careers in the mathematical sciences can click on "Careers and Jobs" at the SIAM website (***www.siam.org***) or proceed directly to* ***www.siam.org/careers****.*

The University of North Carolina at Greensboro

Department of Mathematics and Statistics

The University of North Carolina at Greensboro (UNCG), Department of Mathematics and Statistics seeks applications for a tenure-track assistant professor or tenured associate professor position in computational partial differential equations with a preferred start date of August 1, 2017. Tenure at the associate professor rank may be offered depending on the selected candidate's qualifications. Candidates must hold a Ph.D. in mathematics or a closely related discipline. Competitive applicants will have research expertise that strengthens our Ph.D. program in computational mathematics. Successful applicants will be expected to excel in teaching, maintain a vigorous research program, seek external research funding, contribute to the interdisciplinary mission at UNCG, and educate a diverse group of undergraduate and graduate students from various backgrounds. Application materials should be submitted electronically at http://jobsearch.uncg.edu by clicking on "Faculty" and going to position #999153. Alternately, go directly to https://jobsearch.uncg.edu/postings/8297. Review of applications will begin on April 26, 2017, and will continue until the position is filled. UNCG is especially proud of the diversity of its 43% ethnic minority student body (http://admissions.uncg.edu/discover-about.php). UNCG has been designated as a Minority Serving Institution by the U.S. Department of Education. We seek to attract a diverse applicant pool for this position, especially women and members of minority groups, and we are strongly committed to increasing faculty diversity. UNCG is an EOE AA/M/F/D/V employer.

Ocean Prediction Postdoctoral Positions
Naval Research Laboratory, Stennis Space Center, MS

The Naval Research Laboratory is seeking postdoctoral researchers to push forward the frontiers of ocean forecasting. The work covers a wide scope of physics including surface waves, thermohaline circulation, nearshore circulation, and ocean/atmosphere coupling from global to nearshore scales. This challenging work includes processing and analysis of satellite and in water observations, construction of numerical model systems on high performance computing systems and assimilation for predicting the ocean environment. For a quick overview of some of the research work within the NRL oceanography division at Stennis Space Center, visit the web site: **https://www7320.nrlssc.navy.mil/pubs.php**

U.S. NAVAL RESEARCH LABORATORY

Applicants must be a US citizen or permanent resident at time of application.
Applications will be accepted until positions are filled.
Please e-mail a resume and description of research interests:
Gregg Jacobs:
gregg.jacobs@nrlssc.navy.mil

Deep, Deep Trouble

Deep Learning's Impact on Image Processing, Mathematics, and Humanity

By Michael Elad

I am really confused. I keep changing my opinion on a daily basis, and I cannot seem to settle on one solid view of this puzzle. No, I am not talking about world politics or the current U.S. president, but rather something far more critical to humankind, and more specifically to our existence and work as engineers and researchers. I am talking about…**deep learning**.

While you might find the above statement rather bombastic and overstated, deep learning indeed raises several critical questions we must address. In the following paragraphs, I hope to expose one key conflict related to the emergence of this field, which is relevant to researchers in the image processing community.

First, a few words about deep learning to put our discussion into perspective. Neural networks have been around for decades, proposing a universal learning mechanism that could, in principle, fit to any learnable data source. In its feed-forward architecture, layers of perceptrons—also referred to as neurons—first perform weighted averaging of their inputs, followed by nonlinearities such as a sigmoid or rectified-linear curves. One can train this surprisingly simple system to fit a given input set to its desired output, serving various supervised regression and classification problems.

All of this may sound great, but unfortunately this concept did not take off in the 1980s and 90s — it simply did not provide a sufficiently competitive performance. In addition, the emergence of support vector machines for learning tasks, accompanied by solid theoretical foundations and a convex optimization formulation, seemed to be the last nail in the coffin. Eventually, neural networks entered a long hibernation period. Only a few persistent researchers—Yann LeCun (New York Univerity and Facebook), Geoffrey Hinton (University of Toronto), Yoshua Bengio (University of Montreal), and Jürgen Schmidhuber (Dalle Molle Institute for Artificial Intelligence Research)—stayed in this arena, insisting on trying to convince this seemingly doomed method to behave better. Several important architectures, such as convolutional and long short-term memory networks, resulted from their efforts; yet they were still confined to a niche. Then neural networks suddenly came back, and with a vengeance.

A series of papers during the early 2000s suggested the successful application of this architecture, leading to state-of-the-art results in practically any assigned task. Key aspects in these contributions included the following: the use of many network layers, which explains the term "deep learning;" a huge amount of data on which to train; massive computations typically run on computer clusters or graphic processing units; and wise optimization algorithms that employ effective initializations and gradual stochastic gradient learning. Unfortunately, all of these great empirical achievements were obtained with hardly any theoretical understanding of the underlying paradigm. Moreover, the optimization employed in the learning process is highly non-convex and intractable from a theoretical viewpoint.

This application effort began with written digit recognition (see Figure 1), moving slowly and carefully to more challenging visual and speech recognition and natural language processing tasks, and from there on to practically anything that could be cast as a supervised learning task. Companies such as Google, Facebook, and Microsoft quickly realized the potential in this field and invested massive manpower and budget in order to master these tools and exploit them in their products. On the academic front, conferences in signal processing, image processing, and computer vision have become deep learning playgrounds, contributing to a growing dominance of this bread of work.

This history brings us to present day. For the sake of brevity, consider the classic image processing task of denoising — removing noise from an image (see Figure 2). Thousands of papers addressing this fundamental task were written over the years. Researchers developed beautiful and deep mathematical ideas with tools from partial differential equations, such as anisotropic diffusion and total variation, energy minimization viewpoint, adoption of a geometric interpretation of images as manifolds, use of the Beltrami flow, and more. Harmonic analysis and approximation theory have also served the denoising task, leading to major breakthroughs with wavelet theory and sparse representations. Other brilliant ideas included low-rank approximation, non-local means, Bayesian estimation, and robust statistics. We have hence gained vast knowledge in image processing over the past three decades, impacting many other image processing tasks and effectively upgrading this field to be mathematically well-founded.

In 2012, Harold Burger, Christian Schuler, and Stefan Harmeling decided to throw deep learning into this problem. The idea was conceptually quite simple: take a huge set of clean images, add synthetic noise, and then feed them to the learning process that aims to turn a noisy image into its clean version. While the process was tedious, frustrating, and lengthy—tweaking the method's parameters in a search for good performance likely took a long time—the end result was a network that performed better than *any* known image denoising algorithm at that time.

The above is not an isolated story. Today, deep learning treats many other image processing needs, with unsurpassed results. This is true for single image super-resolution, demosaicing, deblurring, segmentation, image annotation, and face recognition, among others.

Should we be happy about this trend? Well, if we are in the business of solving practical problems such as noise removal, the answer must be positive. Right? Therefore, a company seeking such a solution should be satisfied. But what about us scientists? What is the true objective behind the vast effort that we invested in the image denoising problem? Yes, we do aim for effective noise-removal algorithms, but this constitutes a small fraction of our motivation, as we have a much wider and deeper agenda. Researchers in our field aim to understand the data on which we operate. This is done by modeling information in order to decipher its true dimensionality and manifested phenomena. Such models serve denoising and other problems in image processing, but far more than that, they allow identifying new ways to extract knowledge from the data and enable new horizons.

Now back to the main question: should we be pleased about emerging solutions based on deep learning? Is our frustration justified? What is the role of deep learning in imaging science? These questions present themselves when researchers in the community meet at conferences, and the answers are diverse and confusing. The facts speak loudly for themselves; in most cases, deep learning-based solutions lack mathematical elegance and offer very little interpretability of the found solution or understanding of the underlying phenomena. On the positive side, however, the performance obtained is terrific. This is clearly not the school of research we have been taught, and not the kind of science we want to practice. Should we insist on our more rigorous ways, even at the cost of falling behind in terms of output quality? Or should we fight back and seek ways to fuse ideas from deep learning into our more solid foundations?

To further complicate this story, certain deep learning-based contributions bear some elegance that cannot be dismissed. Such is the case with the style-transfer problem, which yielded amazingly beautiful results, and with inversion ideas of learned networks used to synthesize images out of thin air, as Google's Deep Dream project does. A few years ago we did not have the slightest idea how to formulate such complicated tasks; now they are solved formidably as a byproduct of a deep neural network trained for the completely extraneous task of visual classification.

From my personal viewpoint, image processing researchers have mixed feelings of disgust and envy towards this recent trend of deep learning that keeps pushing itself into our court. Some of us have chosen to remain bystanders for now, while others play along and divert their research agendas accordingly. I belong to the latter group, with some restrictions. In my opinion, it is impossible to imagine that this wave will pass without a marked influence on our field. Thus, I allow deep learning to influence my research team's thoughts and actions, but we continue to insist on seeking mathematical elegance and a clear understanding of the ideas we develop. Time will tell if we are aiming for the impossible.

Briefly circling back to my opening statement on deep learning's massive impact on humankind, human lives will likely be very different several decades into the future. Humanoid robots and intelligence systems might surround us and influence many of our activities, employment and jobs may be things of the past, and relationships between people will probably change drastically. To put it bluntly, your grandchild is likely to have a robot spouse. And here is the punch line: much of the technology behind this bizarre future is likely to emerge from deep learning and its descendant fields.

While this technology progresses rapidly, we haven't stopped to think if this is the future we want for ourselves. The curiosity and tremendous talent of engineers and researchers is driving us towards this future, as do companies that see profit as their main goal. How is it that we rarely engage in discussion about regulating or controlling this progress and guiding it towards a desired future? This is a matter for a different article.

What are your thoughts on the impact of deep learning on image processing — and humanity? Share your feedback by sending us a letter to the editor or blog post at sinews@siam.org, or visit the online version of this article at sinews.siam.org/Current-Issue and post a comment.

Acknowledgments: The author would like to thank Alex Bronstein and Ron Kimmel for valuable comments, which helped tune the message of this article.

References

[1] Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. *Proc. IEEE Int. Conf. Image Process., 1*, I-313-I-316.

Michael Elad is a professor in the Computer Science Department at the Technion – Israel Institute of Technology. He is editor-in-chief of the SIAM Journal on Imaging Sciences.

Figure 2. *A denoising example.* **Left.** *An original image (public domain).* **Middle.** *Image contaminated by additive Gaussian noise of STD=100.* **Right.** *The denoising outcome obtained by one of the leading algorithms — the BM3D [1]. Image credit: Michael Elad.*

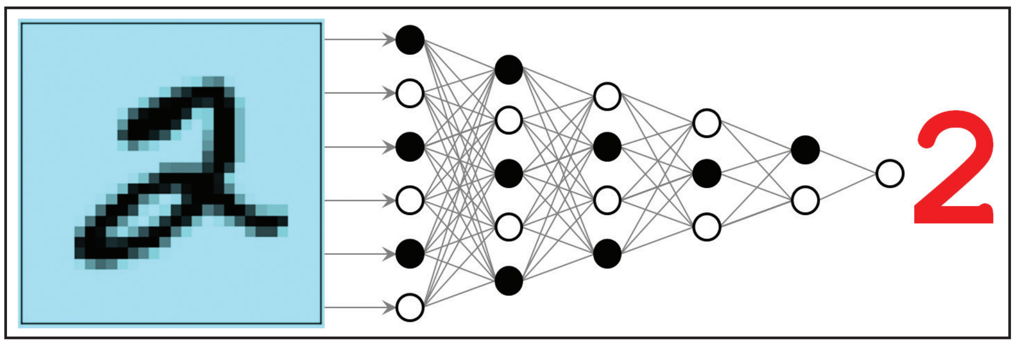


Figure 1. *Neural networks have shown great potential, first in character recognition and subsequently in many other tasks. Image credit: Michael Elad.*