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

siam news

Volume 52/ Issue 9
November 2019

Mathematics + Infrared Technology = New Diagnostic Technique for Autism

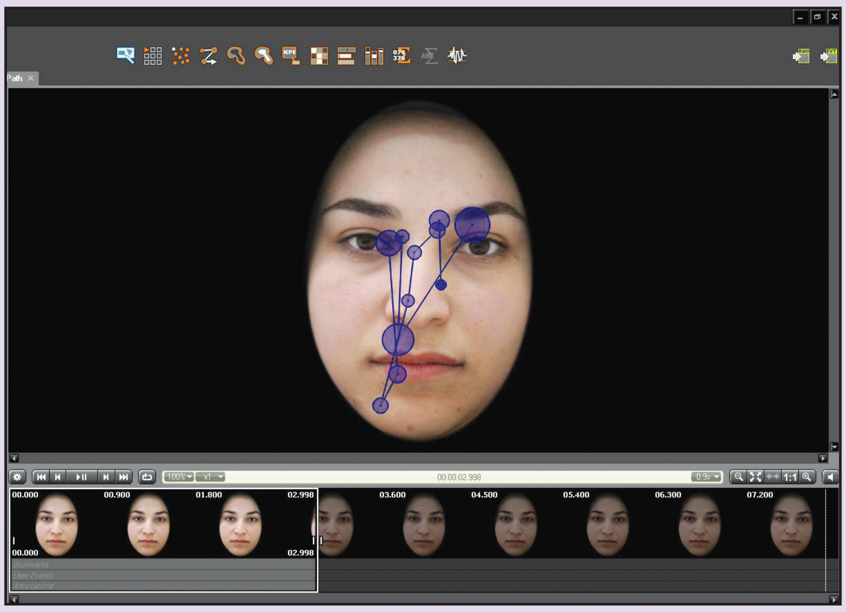


Figure 1. *Researchers can use network analysis to differentiate visual exploration in children with autism spectrum disorder from that in neuro-typical individuals. Photographs of faces on a screen are integrated into an eye-tracking system that interprets and identifies areas of interest on which participants focus their gaze. Figure courtesy of Negar Sammaknejad.*

In an article on page 2, Anita Layton and Mehrshad Sadria present a novel technique that employs network analysis to identify eye-gaze patterns of children with autism spectrum disorder. Such analysis encourages early detection, intervention, and treatment of the disease.

Carbon Cycle Catastrophes: A Dynamical Systems Perspective

By Daniel H. Rothman

Earth's carbon cycles between photosynthesis—which converts carbon dioxide (CO_2) to organic carbon—and respiration, the metabolic processes that transform organic carbon to CO_2 [1, 3]. Since pre-industrial times, human activities have caused atmospheric CO_2 levels to rise by nearly 50 percent, while roughly half as much CO_2 has entered the oceans. While there is considerable interest in predicting climatic responses to these changes, a larger question concerns consequences for the Earth system as a whole — for interactions of life and the environment, in addition to climate. The geologic record indicates that each of the five great mass extinction events of the last 540 million years is associated with significant disruptions of the carbon cycle, as are periods of environmental change unrelated to mass extinction. Although the causes of these events remain ambiguous, the disruptions themselves likely represent a qualitative change in the dynamics of the Earth system. Here I demonstrate how dynamical systems theory helps researchers understand these events.

Observational data sets the stage [4]. Figure 1a (on page 4) displays a geochemical time series containing two disruptions that occurred about 54 million years ago. The sharp downward pulses represent increases Δm in the mass m of inorganic carbon in the ocean. The ratio $\Delta m/m$ can be estimated from the amplitude and duration τ of the downswing. Figure 1b (on page 4) depicts results from 31 events over the last 540 million years. Roughly half of the events lie near the straight line. These *characteristic events* share a similar specific rate r_c in the relation $\Delta m/m = r_c \tau$. Although r_c is biogeochemically significant [4], it may also be dynamically significant because it appears to separate four of the five mass extinction events from nearly all other disruptions.

Carbon cycle disruptions are usually interpreted as proportionate responses to perturbations, such as enhanced emissions of volcanic CO_2, extraordinary releases of methane, or changes in the rates at which organic carbon is sequestered in rock. But such a variety of stressors would seem unlikely to exhibit a common specific rate.

See **Carbon Cycle** *on page 4*

Past and Present of Variational Fracture

By Blaise Bourdin and Gilles A. Francfort

In 1920, Alan Griffith laid the foundations of brittle fracture [10]. His basic postulate was simple; the current state of a crack in an otherwise-elastic material is a cost-benefit analysis between (i) the increment of elastic energy δW_e recovered through the crack's putative infinitesimal advance from its current state, and (ii) the increment of surface energy δW_e spent to achieve such an extension. In short, Griffith's premise was a local minimality principle for the sum $\mathcal{E} := W_e + W_s$ of the elastic and surface energies. He complemented this with the intuitive idea that the surface energy W_s should be proportional to the number of broken bonds, and hence to the length (in two dimensions) or surface area (in three dimensions) of the crack. In two dimensions, $W_s = G_c \ell$, where ℓ denotes the crack length and G_c (the critical energy release rate) is a macroscopic material property.

Griffith's next of kin soon focused on a mere statement of first-order optimality vis-à-vis ℓ for a preordained crack path in two dimensions, i.e., $G := -\partial W_e / \partial \ell \leq G_c$. Because of the connection that George R. Irwin later made between G and the coefficients in front of the singular part of the kinematic field at the crack tip [11], computation of G became the fracture mantra and the sole object of mechanical desire. Three decades later, the research community had forgotten the original variational attitude displayed by its forebearer.

Griffith had remained silent on the topic of the crack path. Path identification was proposed at a cost: the introduction of ad-hoc criteria both for smooth crack paths and for a sudden change of direction (kink) in the crack. With further refinements, this became the field of linear elastic fracture mechanics (LEFM), which held its ground from an engineering standpoint.

The Variational Stance

The variational viewpoint reemerged in 1998 with inspiration from the work of the Ennio De Giorgi school on image segmentation [9]. The idea was to combine minimality with evolution through an energy conservation statement.

Non-interpenetration notwithstanding, the fracture energy associated with a displacement field $u : \Omega \mapsto \mathbb{R}^2$ and a crack Γ in a homogeneous material occupying a two-dimensional domain Ω is

$$\mathcal{E}(u,\Gamma) := W_e(u,\Gamma) + W_s(\Gamma) =$$

$$\int_{\Omega\setminus\Gamma} \frac{1}{2} \mathrm{A}\mathrm{e}(u)\cdot \mathrm{e}(u)\,dx + \int_{\Omega\setminus\Gamma} G_c \mathcal{H}^1(\Gamma),$$

where A denotes the elasticity of the material $\mathrm{e}(u) := \frac{\nabla u + \nabla u^t}{2}$ and $\mathcal{H}^1$ is the one-dimensional Hausdorff measure.

Throughout this evolution, the only transfer of energy is an exchange between elastic and surface energies. Minimality is now global; the cracks are simply sets of co-dimension 1 and finite length (or area). They are only constrained by their past, and add-cracks at each time are topologically free. Cracks are free agents empowered to choose their future paths without a priori imposition of an external criterion.

For scalar-valued kinematic fields—as in the antiplane shear case—existence proofs of a well-posed evolution ensued; this first occurred for closed, connected cracks [7] in a topological setting, then for general cracks [8] in the weak *SBV* setting of De Giorgi and Luigi Ambrosio for the Mumford-Shah image segmentation problem.[1]

A variational phase-field model for fracture [3] soon followed as an adaptation of the Ambrosio-Tortorelli approximation of the Mumford-Shah functional [1] in the sense of Γ-convergence. The corresponding functional is

See **Variational Fracture** *on page 3*

[1] The generalization to elasticity proved challenging and the three-dimensional case is still pending.

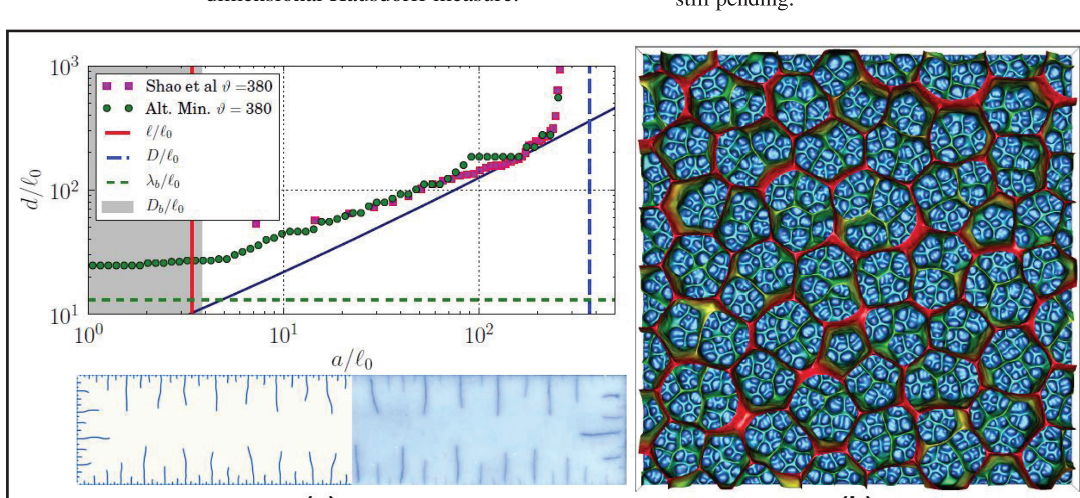


Figure 1. *Phase-field simulation of a thermal shock [4].* **1a.** *Crack spacing versus distance to edge, and crack geometry in numerical simulation versus experiments.* **1b.** *Complex crazing-like fracture pattern growing from the edge toward the inside of the domain in a large-scale, three-dimensional numerical simulation. The level line $\alpha = 0.95$ of the phase field is colored by the distance to the exposed surface; blue is toward the exposed face and red is inside the material. Figure courtesy of [4].*

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 52/ Issue 9/ November 2019

Mathematics + Infrared Technology = New Diagnostic Technique for Autism

By Anita Layton and Mehrshad Sadria

Chances are you know someone with concerns about their toddler's behavior. Their child may be shy and sweet, but have an unpredictable attitude. He or she might throw the worst temper tantrums, sometimes kicking and screaming inconsolably for an hour. The smallest changes in routine may throw him or her off. It is probably impossible to reason with the child when he or she is upset. Does this behavior simply indicate a bad case of the "terrible twos"? Should the child be allowed time to grow out of this phase? Or are these signs of autism spectrum disorder (ASD)?

ASD is a neurodevelopmental disorder that affects approximately one to two percent of the population. This means that a school bus full of children will typically have one or two kids with ASD. Symptoms usually appear in the first two years of life and impair a child's ability to function socially. Although current treatments vary, most interventions focus on managing behavior and improving social and communication skills with the hope that the child will one day become an independent adult. Because the capacity for change is greater at a young age, early diagnosis and intervention ensure the best possible outcomes.

Mathematics as the New Microscope

Given the many benefits of timely detection and treatment, we are keen to help doctors more quickly and accurately diagnose ASD in children. Our group uses mathematics as a microscope to understand biology and medicine [1]. We build computer models to simulate the effects of various drugs [2-3] and apply mathematical techniques to analyze clinical data.

How can we utlize math to generate a new autism detection tool? We believe that mathematics can objectively distinguish behaviors of children with ASD from those of their neuro-typical counterparts. For instance, we know that visual exploration in patients with ASD is different from that in neuro-typical individuals. This knowledge motivated us to create a new ASD detection technique—based on network analysis—that distinguishes varied eye-gaze patterns [4].

Network-based Analysis for Eye-gaze Patterns

To develop our novel method, we first evaluated 40 four- or five-year-old children. Half of them had ASD, while the others were neuro-typical. We showed each participant 44 photographs of faces on a screen, which were integrated into an eye-tracking system (see Figure 1, on page 1). The infrared device uses emission and reflection of waves from the iris to interpret and identify locations on the stimuli at which each child was looking.

We then separated the images into seven key features, or areas of interest (AOI), on which participants focused their gaze: under the right eye, on the right eye, under the left eye, on the left eye, on the nose, on the mouth, and on other parts of the screen. We considered both a static measure—how much time each participant spent looking at each feature—and dynamic measures — how they moved their eyes and scanned the faces [4]. Children with ASD focus on and scan faces differently. For instance, when looking at a person's face, neuro-typical children focus more on the eyes and those with ASD focus more on the mouth. This static measure is called "fixation time." And when shifting their focus from someone's eyes to their chin, neuro-typical children generally move their eyes more quickly and via a different path than children with ASD.

To investigate how ASD and neuro-typical children explore facial features in dissimilar ways, we employ the centrality concept from network analysis. Here, each of the seven AOIs acts as a node in our network model. Every saccadic transition between two AOIs yields a link connecting the two respective nodes. One can apply measures of centrality to assess the importance of each node. We computed the subsequent four centrality measures.

(i) Degree centrality—arguably the simplest conceptually—depicts the number of links associated or connected with a given node.

(ii) Betweenness centrality is the number of shortest paths between two additional nodes that pass through a given node. One can compute the betweenness of node k in $G=(V;E)$ as follows:

$$B_k = \frac{\Sigma_{i,j\neq k} P_k(i,j) / P(i,j)}{\begin{pmatrix} N-1 \\ 2 \end{pmatrix}},$$

where N is the number of nodes in the graph and $P(i,j)$ denotes the number of shortest paths between nodes i and j. Among these shortest paths, $P_k(i,j)$ indicates how many pass through node k.

(iii) Closeness centrality is another measure of a given node's importance. One can calculate it from the reciprocal of the sum of the shortest paths' length between the node and all other nodes in the graph. Closeness centrality quantifies the extent to which the network may be congested between a given pair of nodes:

$$C_k = \frac{N-1}{\Sigma_j d_{i,k}},$$

where $d_{i,k}$ is the shortest path between nodes i and k.

(iv) Eigenvector centrality assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. Specifically, the eigenvector centrality of node k—denoted E_k—is proportional to the weighted sum of the eigenvector centrality of the nodes to which it connects:

$$E_k = X_k^{\max} = \frac{1}{\lambda^{\max}} \sum_j a_{k,j} X_j^{\max},$$

where $\lambda^{\max}$ is the largest eigenvalue related to the transition matrix $(a_{i,j})$ and $X^{\max}$ is the associated eigenvector.

A More Child-friendly Test for Early Autism Detection

Why do we want a new diagnostic technique for ASD? The hope is that a more child-friendly and improved diagnostic tool would facilitate early ASD detection in children. There is currently no definitive laboratory test or biomarker for the disorder. Assessment often includes a medical and neurological examination; an in-depth questionnaire about the child's family history, behavior, and development; or an evaluation from a psychologist. In addition to being expensive, these diagnostic approaches are not toddler-friendly. Children will no doubt find it much easier to stare at faces than fill out questionnaires or undergo a psychological evaluation.

We hope that our network-based analysis will make the diagnostic process less stressful. Parents who suspect ASD in their toddlers can seek screening with our technology and have their children scan a few faces. If the results point to a high likelihood of ASD, the child can then undergo a more involved clinical appraisal.

Our group is mindful of the many benefits that accompany early ASD diagnosis and subsequent early intervention. Timely detection is associated with reduced parental stress, and prompt intervention is critical to achieving positive outcomes over time. Treatments implemented before age four are associated with significant gains in cognition, language, and adaptive behavior. Studies have similarly linked early interventions in ASD with improvements in daily living skills and social conduct.

By using mathematics to remove some of the barriers that inhibit early diagnosis, we hope that more children with ASD can receive timely intervention and attain a better quality of life and a higher degree of long-term independence.

References

[1] Layton, A.T. (2017). A new microscope for the kidney: mathematics. *Am. J. Physiol. Renal Physiol., 312*, F671-F672.

[2] Layton, A.T., & Vallon, V. (2018). SGLT2 inhibition in a kidney with reduced nephron number: modeling and analysis of solute transport and metabolism. *Am. J. Physiol. Renal Physiol.*, *314*, F969-F984.

[3] Leete, J., & Layton, A.T. (2019). Sex-specific long-term blood pressure regulation: modeling and analysis. *Comp. Bio. Med., 104*, 139-148.

[4] Sadria, M., Karimi, S., & Layton, A.T. (2019). Network centrality analysis of eye-gaze data in autism spectrum disorder. *Comput. Bio. Med., 111*, 103332.

Anita Layton is the Canada 150 Research Chair in Mathematical Biology and Medicine and a professor of applied mathematics, pharmacy, and biology at the University of Waterloo. She is also the associate dean (Research and International) for the Faculty of Mathematics at Waterloo. Mehrshad Sadria is a MMath candidate in the Department of Applied Mathematics at the University of Waterloo. He received a bachelor's degree in physics from Shahid Beheshti University (Tehran, Iran) in 2017.

siam news

ISSN 1557–9573.

Variational Fracture

Continued from page 1

$$\mathcal{E}_\varepsilon(u,\alpha) := \int_\Omega \frac{1}{2}\big(a(\alpha)+\eta_\varepsilon\big)\mathrm{Ae}(u)\cdot \mathrm{e}(u)\,dx$$

$$+\frac{G_c}{4c_w}\int_\Omega \left(\frac{w(\alpha)}{\varepsilon}+\varepsilon\,|\nabla\alpha|^2\right)dx,$$

with $\alpha : \Omega \mapsto [0,1]$ as the phase-field variable, ε and $\eta_\varepsilon \ll \varepsilon$ as regularization parameters, and a and w as two continuous strictly-monotonic functions, such that $w(0)=a(1)=0$, $w(1)=a(0)=1$, and $c_w := \int_0^1 \sqrt{w(s)}\,ds$. The approximate evolution proved numerically palatable and offered practitioners a computational tool for complex crack evolutions.

The mathematical activity surrounding variational and phase-field models of fracture initially went unnoticed by the larger fracture community, even as it flourished in the mathematical literature. Meanwhile, a growing interest in complex fracture problems was challenging practitioners to put forth tools that exceeded the capabilities of LEFM. A decade after the introduction of the aforementioned variational model, the fracture community experienced a sudden outburst of phase-field models (and paternity claims). The phase-field approach to fracture was on its way to its current quasi-hegemonic status.

Knowns and Unknowns of the Variational Model

The variational approach dispels the misbegotten notion that crack path prediction is subordinate to specific criteria. In particular, local stability combined with energy preservation adjudicates the rivalry between the two main kinking standards: maximality of the elastic energy release rate versus local symmetry of the kinematic fields after kinking. Both criteria must be satisfied simultaneously. Since they contradict each other [2], the only possible outcome is that kinks do not exist. Griffith's theory, spawned from local stability, actually disavows the addition of the ad-hoc kinking criteria meant to supplement it [5]. On the computational front, phase-field approximations have repeatedly demonstrated their ability to quantitatively predict fracture evolution [4].

Global minimality of the crack state at each time was key to mathematization of the evolution problem. In physics, however, any minimality postulate is technical convenience rather than the manifestation of a known physical imperative. In this case, global minimality in concert with free crack path has two serious drawbacks. First, it behaves hideously when the sample is subject to a force load — say $f(x,t)$ in its interior. Indeed, one should account for the work of $f(x,t)$ in such a case, thereby adding $-\int_\Omega f.u\,dx$ to $\mathcal{E}$. However, doing so prevents the realization of minimality because cutting off the support of $f(x,t)$ and moving it away from Ω drives that linear term to $-\infty$ while requiring only the price of that cut in surface energy and canceling the elastic energy altogether.

Because global minimality scans the entire energetic landscape, it enables situations wherein a steadily-growing crack would leverage information on distant materials properties or geometric features of which it has no possible knowledge. Therefore, abandoning global minimality in favor of a more local criterion would seem reasonable. Unfortunately, doing so would not cure either ailment: force loads would still favor immediate removal of their support, while initiation would be impossible for any kind of decent topology [6]. In addition, the link between phase-field and variational models would be severed, as Γ-convergence does not generally imply convergence of local minimizers.

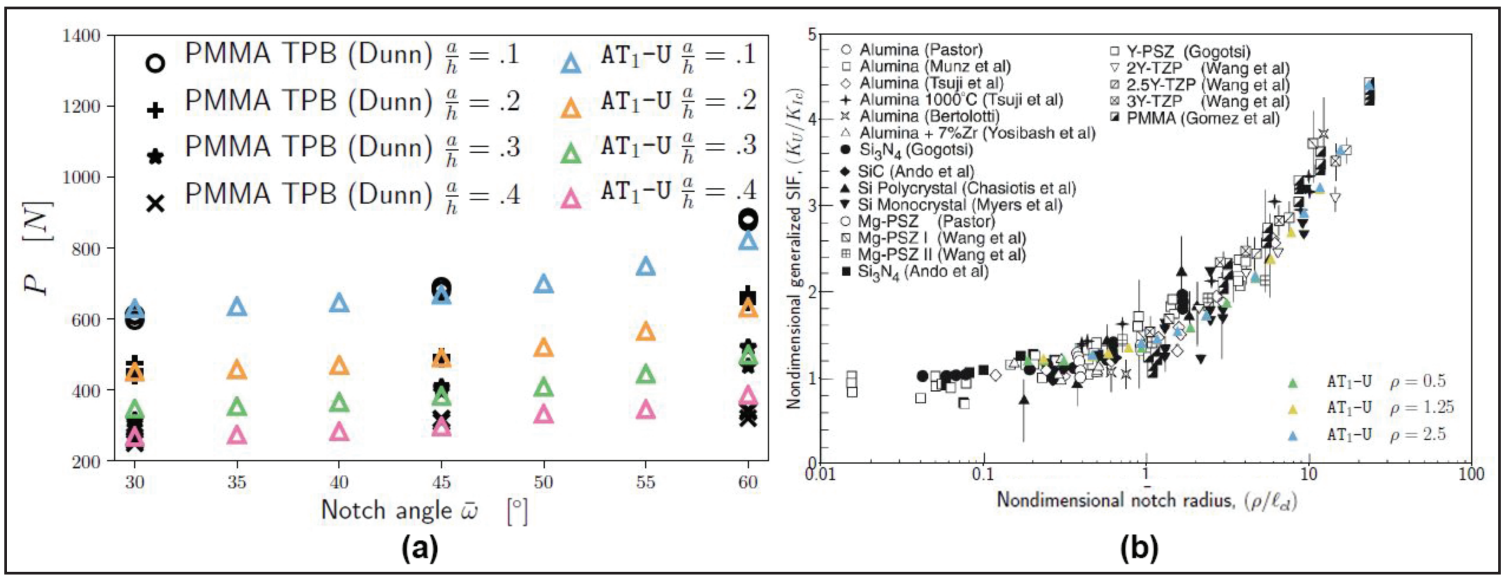


Figure 2. *Crack nucleation in phase-field simulations (colored triangles) and experiments [12].* **2a.** *Critical fracture load in a three-point bending experiment, with a V-notch as a function of the notch-relative depth a/h and opening half angle (black symbols) $\bar{\omega}$.* **2b.** *Phase-field computation of the generalized stress intensity factor near a U-notch, compared to experiments in a wide range of materials. Figure courtesy of [12].*

Phase Field as an Autonomous Agent

In recent years, researchers have used phase-field models for an increasingly wide range of applications, including coupled problems like thermal cracking or hydraulic fracturing. They have also contemplated extensions to ductile and dynamic fracture propagation. These models are mostly ad-hoc, and mathematical understanding is partial at best. Yet conclusive quantitative comparisons with experiments, such as that in Figure 1 (on page 1), cannot be ignored.

As mentioned above, the phase-field model finds its justification as an approximation of fracture through Γ-convergence, which is only concerned with global minimization. But global minimization is unrealistic. Moreover, there are no foolproof algorithms for computing global minima because $\mathcal{E}_\varepsilon$ is non-convex. Over the past 20 years, researchers have strived to incorporate a modicum of minimality into the computation of the phase-field evolution, with the understanding that doing so may lead them far from the fracture model [12]. The alternative is to view the phase-field model as the physical parent, thereby equating brittle fracture with a nonlocal damage gradient model — a radical step from a mechanician's standpoint.

To make things more complex, it is possible within the context of the phase-field approximation to devise a scheme that—by judiciously fixing the length scale—delivers the correct initiation sequence when quantitatively tested against experiments (see Figure 2). It is both remarkable and utterly baffling that well-devised phase-field evolutions provide such accurate approximations of fracture evolutions, from crack nucleation to sample failure.

The variational view of fracture has delivered a mathematically-consistent theory that quantitatively reproduces many experimental results. Yet its ability to do so is shrouded in mystery. The initial impetus to recast brittle fracture in a variational framework was motivated by the relevant advances in mathematics, with mechanics arguably lagging behind. The situation has since reversed; the past few years have witnessed the resurgence of mechanical considerations, especially in terms of phase-field models. The mathematical tools required for rigorous investigation of topics like initiation, local minimality, or inertial effects are still in their infancy.

References

[1] Ambrosio, L., & Tortorelli, V.M. (1990). Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Comm. Pure Appl. Math., 43*(8), 999-1036.

[2] Amestoy, M., & Leblond, J.-B (1989). Crack paths in plane situation – II. Detailed form of the expansion of the stress intensity factors. *Int. J. Solids Struct., 29*(4), 465-501.

[3] Bourdin, B., Francfort, G.A., & Marigo, J.-J. (2000). Numerical experiments in revisited brittle fracture. *J. Mech. Phys. Solids, 48*, 797-826.

[4] Bourdin, B., Marigo, J.-J., Maurini, C., & Sicsic, P. (2014). Morphogenesis and propagation of complex cracks induced by thermal shocks. *Phys. Rev. Lett., 112*(1), 014301.

[5] Chambolle, A., Francfort, G.A., & Marigo, J.-J. (2009). When and how do cracks propagate? *J. Mech. Phys. Solids, 57*(9), 1614-1622.

[6] Chambolle, A., Giacomini, A., & Ponsiglione, M. (2008). Crack initiation in brittle materials. *Arch. Ration. Mech. Anal., 188*(2), 309-349.

[7] Dal Maso, G., & Toader, R. (2002). A model for the quasi-static growth of brittle fractures: Existence and approximation results. *Arch. Ration. Mech. Anal., 162*(2), 101-135.

[8] Francfort, G.A., & Larsen, C. (2003). Existence and convergence for quasi-static evolution in brittle fracture. *Comm. Pure Appl. Math., 56*(10), 1465-1500.

[9] Francfort, G.A., & Marigo, J.-J. (1998). Revisiting brittle fracture as an energy minimization problem. *J. Mech. Phys. Solids, 46*(8), 1319-1342.

[10] Griffith, A.A. (1921). The phenomena of rupture and flow in solids. *Phil. Trans. Roy. Soc. Lond., 221*(582-593), 163-198.

[11] Irwin, G.R. (1958). Fracture. In *Handbuch der Physik, herausgegeben von S. Flügge. Bd. 6. Elastizität und Plastizität* (pp. 551-590). Berlin: Springer-Verlag.

[12] Tanné, E., Li, T., Bourdin, B., Marigo, J.-J., & Maurini, C. (2018). Crack nucleation in variational phase-field models of brittle fracture. *J. Mech. Phys. Solids, 110*, 80-99.

Blaise Bourdin is the A.K. & Shirley Barton Professor in the Department of Mathematics at Louisiana State University. Gilles Francfort is a professor in the Department of Mathematics at Université Paris 13.

Carbon Cycle

Continued from page 1

Characteristic events may instead reflect the intrinsic dynamics of the carbon cycle.

A simple model of the marine carbon cycle illustrates how this could work [5]. We consider the upper ocean to be a well-mixed chemical reactor, open to an incoming flux j_i of dissolved calcium carbonate from rivers and a maximum outgoing flux bj_i to sediments. Dissolved inorganic carbon in the ocean takes several forms; it suffices to consider only two. We track the concentration of total dissolved inorganic carbon (carbonate and bicarbonate ions in addition to CO_2), denoted by w, and the concentration of carbonate ions (CO_3^{2-}), denoted by c. We also allow for external input of CO_2 at rate νj_i.

One important feedback concerns the outgoing flux. Above a critical concentration c_p, carbonate minerals are preserved in sediments; below c_p, they dissolve. Averaged over the oceans, the transition is smooth rather than sharp. We therefore assume that the outflux is proportional to a sigmoidal function $s(c,c_p)$ that grows from zero to one as c increases to values well above c_p.

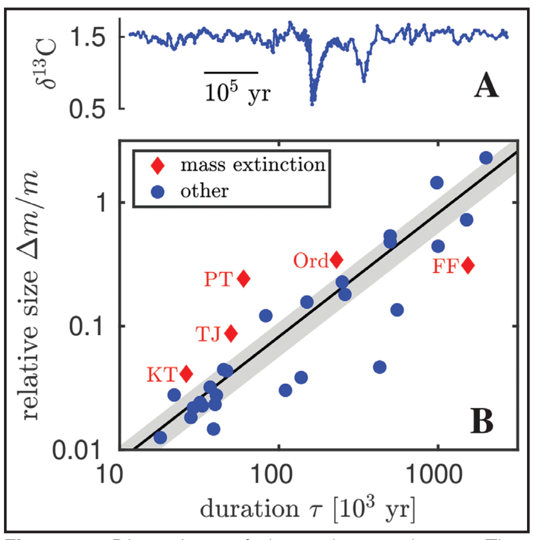


Figure 1. *Disruptions of the carbon cycle.* **1a.** *Time series of the isotopic composition of carbonate carbon ($\delta^{13}C$, the relative enrichment of ^{13}C compared to ^{12}C, expressed as "per mil") during the Eocene period, about 54 million years ago. The two abrupt downswings correspond to increases Δm in the mass m of dissolved inorganic carbon in the oceans.* **1b.** *The relative size and duration of 31 global disruptions (including those in 1a) in the last 540 million years. The duration τ is the time over which m grows. The labeled events are associated with the end-Cretaceous (KT), end-Triassic (TJ), end-Permian (PT), end-Ordovician (Ord), and Frasnian-Famennian (FF) mass extinctions. Figure 1a adapted from [7], 1b adapted from [4].*

Another feedback concerns the carbonate system's interaction with planktonic organisms. When CO_2 is added to seawater, some of it combines with carbonate ions to form bicarbonate ions; this lowers the concentration c of carbonate ions. If c becomes sufficiently small—less than a concentration c_x—planktonic organisms with calcium carbonate shells fare poorly. Their shells normally provide "ballast" that causes them to sink to the deep sea with detrital organic carbon. But if fewer calcifying organisms exist in the upper ocean, less organic carbon sinks to the seafloor and more is respired to CO_2 near the sea surface, causing upper ocean CO_2 levels to increase even further. We express this process by a constant θ times a sigmoidal function $\bar{s}(c,c_x)$, where $\bar{s} = 1 - s$. Finally, we assume that the concentration of total dissolved inorganic carbon tends to relax toward a concentration w_0 at a timescale τ_w. Today, $\tau_w \sim 10^4$ years. These considerations yield the *carbon cycle model*:

$$\dot{c}/f(c) = j_i[1 - bs(c,c_p) - \theta\bar{s}(c,c_x) - \nu] + (w - w_0)/\tau_w \quad (1)$$

$$\dot{w} = j_i[1 - bs(c,c_p) + \theta\bar{s}(c,c_x) + \nu] - (w - w_0)/\tau_w. \quad (2)$$

The function $f(c)$ approximates the manner in which the carbonate system "buffers" the addition or removal of chemical species.

Depending on the parameters, the model exhibits a stable fixed point, a stable limit cycle, or both. Here we focus on the stable fixed point's response to perturbation when the fixed point is near the bistable region. We imagine that the system is initially prepared with the CO_2 injection parameter $\nu = 0$, but we set $\nu = \nu_0 > 0$ for all times $t \geq 0$. Figures 2a and 2b show that the system spirals toward a new steady state. Yet when ν_0 is larger than a threshold ν_c, the system undergoes a large excitation before approaching the new fixed point (see Figures 2c and 2d). The model is therefore *excitable*, similar to models of action potentials in neurons [2]. The excitation corresponds to transient ocean acidification, since the addition of CO_2 reduces not only c but also the pH.

Because the size and timescale of excitations are properties of the system rather than its perturbation, an excitable carbon cycle could explain several features of the data in Figure 1b [5]. Characteristic events, which fall close to the straight line, would represent near-threshold excitations. The mass extinction events above the line would be associated with perturbations that significantly exceed the threshold. Events below the line may simply represent slow quasistatic change.

In the real carbon cycle, initiation of an excitation by CO_2 injection would likely last for only a finite time t_i, e.g., $\nu = \nu_0$ for $0 \leq t \leq t_i$. In the carbon cycle model, the excitation threshold ν_c is independent of t_i when t_i exceeds approximately one damping time τ_w. But if $t_i \ll \tau_w$, perturbations can be damped before they lead to excitation. In this case, the threshold depends on the total mass injected and $\nu_c \propto t_i^{-1}$. Figures 3a and 3b illustrate how this works. Similar phenomena exist in other excitable systems [2, 8].

The flux of CO_2 entering today's oceans is nearly two orders of magnitude greater than the growth rate of characteristic events, which provides a rough estimate of the threshold's upper bound under a step-like injection. Yet the centennial timescale of the modern perturbation is about two orders of magnitude smaller than today's damping timescale. The shorter injection timescale cancels out the stronger injection rate, and the modern influx lies near the hypothesized upper bound rather than greatly exceeding it.

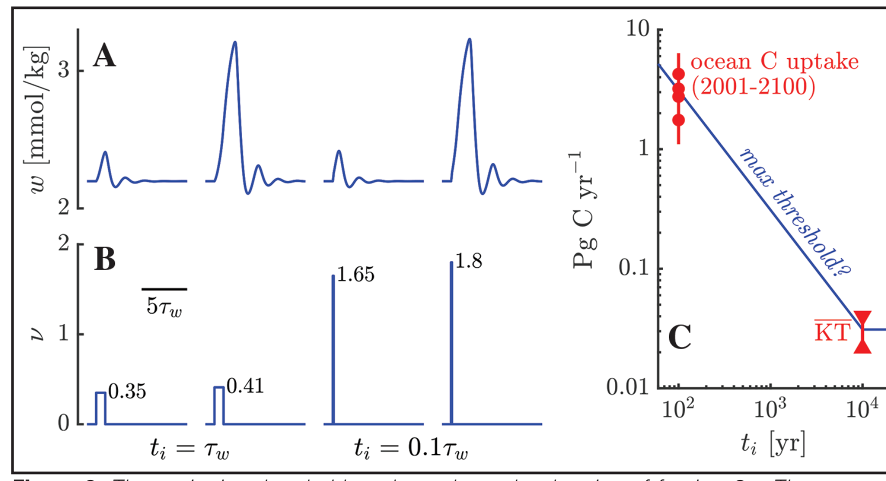


Figure 3. *The excitation threshold ν_c depends on the duration of forcing.* **3a.** *The response $w(t)$ to CO_2 injection (in* **3b***) at varied dimensionless rates ν over different durations t_i.* **3c.** *Comparison of perturbations of the modern and end-Cretaceous [6] carbon cycles to a hypothesis for the upper bound of the excitation threshold. Four projections of the modern perturbation are depicted [5]. The end-Cretaceous case $\overline{KT}$ represents the upper half of the confidence interval for CO_2 emitted by massive volcanism tens of thousands of years before the end-Cretaceous extinction [6]. Because the excitation threshold scales like t_i^{-1}, both the modern and ancient perturbations are equivalently near it. Figure 3a and 3b courtesy of Daniel Rothman, 3c adapted from [5].*

To understand what this might mean, consider the carbon cycle's behavior before the end-Cretaceous extinction (and the dinosaurs' demise). Although the extinction is widely attributed to a bolide impact, modern geochronological methods reveal a roughly 10^4-year pulse of massive volcanism tens of thousands of years prior to impact [6]. The resulting CO_2 injection occurred at an estimated peak rate that is approximately one percent of the maximum projected 21st-century mean flux to the oceans [5]. Both the modern and end-Cretaceous CO_2 fluxes therefore lie near the threshold's upper bound (see Figure 3c). Massive volcanism associated with the end-Permian and end-Triassic extinction events may be similarly pulsed.

The upshot is twofold. First, modest perturbations of the carbon cycle beyond a threshold may have led to significant disruptions of the ancient Earth system, possibly including mass extinction. Second, today's strong perturbation appears to be near an equivalent threshold. Are we therefore headed for a sixth extinction? The dynamical system of equations (1) and (2) is a toy model; among other limitations, its assumption of a well-mixed ocean neglects possible feedbacks on timescales less than approximately 10^3 years. Yet the hypothesis of excitability is reasonable and the t_i^{-1} scaling law provides something novel and important: a means of rescaling the past to the present. Analysis of alternative models and acquisition of more data will help test these ideas. The risk of catastrophe makes fundamental progress imperative.

Acknowledgments: I thank Constantin Arnscheidt for helpful discussions. This work was supported by NASA astrobiology grant NNA13AA90A and NSF grant 1338810.

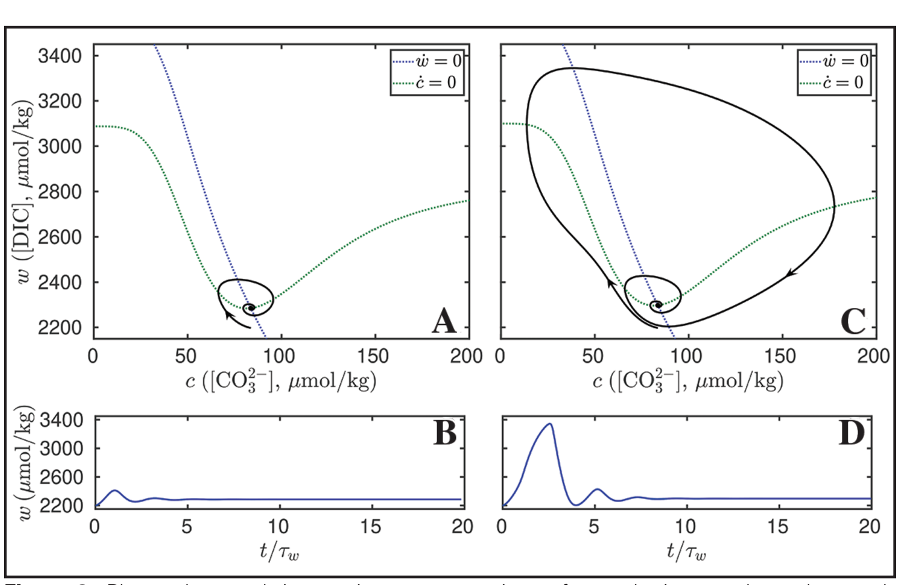


Figure 2. *Phase plane and time series representations of perturbations to the carbon cycle model below (***2a** *and* **2b***) and above (***2c** *and* **2d***) the excitation threshold. Image courtesy of [5].*

References

[1] Archer, D. (2010). *The Global Carbon Cycle*. Princeton, NJ: Princeton University Press.

[2] Izhikevich, E.M. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge, MA: MIT Press.

[3] Rothman, D.H. (2015). Earth's carbon cycle: a mathematical perspective. *Bull. Amer. Math. Soc., 52*, 47-64.

[4] Rothman, D.H. (2017). Thresholds of catastrophe in the Earth system. *Sci. Adv., 3*(9), e1700906.

[5] Rothman, D.H. (2019). Characteristic disruptions of an excitable carbon cycle. *PNAS, 116*, 14813-14822.

[6] Schoene, B., Eddy, M.P., Samperton, K.M., Keller, C.B., Keller, G., Adatte, T., & Khadri, S.F.R. (2019). U-Pb constraints on pulsed eruption of the Deccan Traps across the end-Cretaceous mass extinction. *Science, 363*, 862-866.

[7] Westerhold, T., Röhl, U., Frederichs, T., Agnini, C., Raffi, I., Zachos, J.C., & Wilkens, R.H. (2017). Astronomical calibration of the Ypresian timescale: implications for seafloor spreading rates and the chaotic behavior of the solar system? *Clim. Past, 13*, 1129-1152.

[8] Wieczorek, S., Ashwin, P., Luke, C.M., & Cox, P.M. (2011). Excitability in ramped systems: the compost-bomb instability. *Proc. Roy. Soc. A, 467*(2129), 1243-1269.

Daniel H. Rothman is a professor of geophysics and co-director of the Lorenz Center in the Department of Earth, Atmospheric, and Planetary Sciences at the Massachusetts Institute of Technology.

Changes to SIAM's Child Care Grant Program

SIAM offers child care grants that provide monetary support for those attending select SIAM conferences with small children.

Beginning in 2020, SIAM will be changing the way in which it administers these grants. SIAM will provide a grant amount of \$85 per family per conference day (as opposed to the \$250 total per conference that attendees received in the past). The new policy better aligns with a 2018 study conducted by *Science* magazine's careers department.

Child care services, caregiver travel, and other conference-related expenses incurred in the care of dependent children are eligible for reimbursement if they result directly from the parent's participation in the conference, regardless of whether care is provided at the conference site or at home.

Since their introduction in 2012, SIAM child care grants have been helping working parents balance their careers and family lives. To apply for a child care grant, look for the application portal on the conference website under "Lodging & Support."

Computational Mechanics: Three Decades of Applications and Extensions

By James P. Crutchfield

In part I[1] of this article, published in the October issue of SIAM News, the author provided a thorough introduction to computational mechanics that documented the field's multifaceted origin.

My interest in computational mechanics began as a fascination with mainframe computers in the 1960s and information theory in the 1970s. I worked in Silicon Valley for a number of years, first at IBM (at what was to become its Almaden Research Center) on information storage technology — specifically magnetic bubble devices. This was followed by an assignment at Xerox's Palo Alto Research Center, which at the time was busily inventing our current computing environment of packet-based networks (ethernet), internet protocols, graphical user interfaces, file servers, bitmap displays, mice, and personal workstations. An active member of the Homebrew Computer Club, I hand-built a series of microcomputers: 4-bit, 8-bit, and eventually 16-bit machines. I suggested and helped code the first cellular automaton simulator on a prototype 6502 (8-bit) microcomputer, which ultimately became the Apple I. During this time, I met many people who would later become titans of modern information technology.

As a student at the University of California, Santa Cruz (UCSC), I learned about the mathematics of computers and communication theory directly from information theory pioneer David Huffman, who was well known for his 1950s work on minimal machines — specifically *machine synthesis*. Huffman's pioneering research was an integral part of his discrete mathematics and information theory courses. I learned computer architecture from Harry Huskey, a chief engineer for the first U.S. digital computers (ENIAC and EDVAC) who also taught at UCSC. In short, thinking about computing and its physical substrates went hand in hand with my physics training in statistical mechanics and my mathematics training in dynamical systems theory. This theme drove the bulk of my research on chaotic dynamics.

With this background in mind, I will now address the immediate concerns of nonlinear physics in the 1980s. As computers decreased in size and cost, they became increasingly accessible research tools. In the late 1970s and early 80s, this revolution inspired the burgeoning field of nonlinear dynamics. Computer simulations, unlike abstract existence proofs, allowed us to simply examine and interact with the solutions of complex nonlinear systems. The new tools thus revealed an unexplored universe of exquisitely complex, highly-ramified structures and unpredictable behaviors that had remained relatively abstract throughout most of the 20th century.

Randomness emerged spontaneously — though paradoxically, we knew and had programmed the underlying equations of motion. This presented several deep challenges. What is randomness? Can we quantify it? Can we extract the underlying equations of motion from observations? Was each and every nonlinear system, in the vast space of all systems, going to require its own "theory"?

In the 1970s, the nonlinear physics community identified a target problem—fluid turbulence—to probe these questions and presented a testable hypothesis, specifically the Ruelle-Takens conjecture that strange attractors were the internal mechanism driving the problem. This formalized meteorologist Edward Lorenz's earlier proposal on deterministic nonperiodic

See **Computational Mechanics** *on page 7*

[1] https://sinews.siam.org/Details-Page/dynamics-information-and-organization-the-origins-of-computational-mechanics

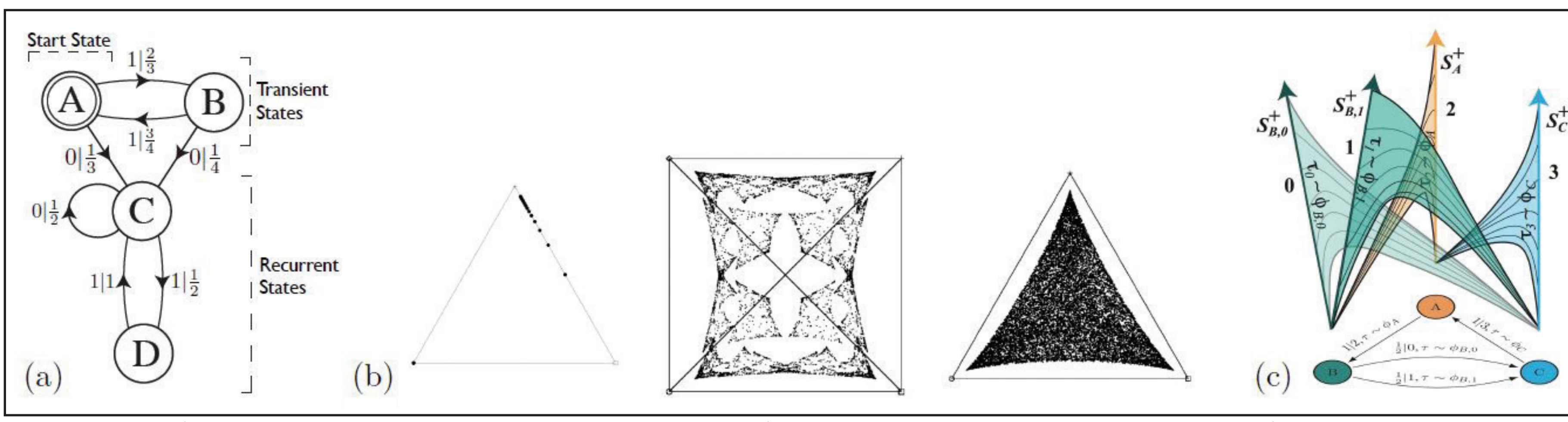


Figure 1. *Varieties of the ϵ-machine.* **1a.** *Finite-state hidden Markov model (HMM).* **1b.** *Infinite mixed state: countable HMM and uncountable HMM (fractal and continuum).* **1c.** *ϵ-machine for discrete-event, continuous-time hidden semi-Markov processes. Figure courtesy of James Crutchfield.*

Preparing Applied Mathematics Students for Careers in Industry

A Recurring Theme at ICIAM 2019 Education Sessions

By Lucas Castle

Applied mathematicians know the importance of recognizing and promoting mathematical innovation as it relates to industry, medicine, and engineering. They realize that interdisciplinary partnerships are vital to solving the world's most complex problems. Thousands of mathematicians and computational scientists from around the globe recently convened in Valencia, Spain, for the 9th International Congress on Industrial and Applied Mathematics, which took place this July. Attendees shared recent mathematical advances, collaborated with peers, and facilitated partnerships between academia and industry.

The need for applied mathematics is abundantly clear, and education is a particularly critical component. Presenters at the various minisymposia on mathematics education sought to answer the following question: How are we leveraging our skills as applied mathematicians to truly train and prepare students? Each talk offered a different perspective on this meaningful, complex problem. Some educators spoke of their innovative, problem-focused courses, while others shared their experiences with exposing students—both at the undergraduate and graduate level—to mathematics' relevance via modeling competitions or industry collaborations. In all cases, students pondered the real-world applications of mathematics both in and out of the classroom.

Inside the Classroom

Undergraduate students often ask me "Why is this useful?" While answering this question is relatively easy, simply listing applications of relevant mathematical concepts still leaves students somewhat unsatisfied. To bypass this issue altogether and make mathematical applications self-evident, many educators are shifting towards problem-centric curriculum design, in which students experience mathematics by tackling relevant and contextualized problems. For example, linear algebra students might develop notions of linear systems by analyzing traffic flow in a series of intersections or studying disease dynamics in epidemic models. Undergraduates in a quantitative reasoning course could design an art gallery—complete with a scale model—to understand the steps involved in the mathematical modeling process. These and other practical experiments effectively demonstrate (rather than simply state) mathematics' applicability in the real world.

Courses that incorporate projects of this nature also stimulate the development of other desirable, non-mathematical skills. Open-ended problems set the stage for collaborative, multidisciplinary experiences in inquiry-based environments. However, solving a problem is only half the battle. Communication of solutions, both within a group and to intended stakeholders, provides opportunities for students to practice writing and speaking skills via a variety of media. A group performing sensitivity analysis on a model of a stocked lake might convey its results by writing a memo to the U.S. Fish and Wildlife Service with a recommended fish harvesting rate for the lake. Alternatively, a small modeling project on international travel could culminate with oral presentations to a "donor" (in this case, the instructor) that request funding for the hypothetical trip. Using mathematics as the vehicle through which students develop these soft skills reinforces the relevancy of the subject in context.

Of course, teaching and learning in this style is not without challenges. These courses necessitate flexibility and require buy-in from both students and instructors. Teachers must be prepared to generate and facilitate open-ended problems that have no "right" answer. Assessment is also more complicated; grading a particular assignment that yields a variety of solutions, deciding whether traditional hour-long exams fit the material, and balancing rubrics between mathematical concepts and soft skills (such as writing, communicating information or data effectively, etc.) pose difficulties. That said, the rewards far outweigh the obstacles. Contextualized mathematics enhances student attention, information retention, and reflective learning. Students find material in this framework more applicable to their lives and future careers, as the mathematics in question actually matters to them. Furthermore, undergraduates in courses that challenge them with "real problems" tend to outperform their peers in classes that utilize more traditional teaching structures.

CAREERS IN MATHEMATICAL SCIENCES

Beyond the Classroom

Authentic classroom problems especially benefit students in classes that are taught from a more traditional perspective (like service courses and lower-division math courses). Students taking advanced courses are also likely to encounter at least a few of these types of questions. But no matter what, the classroom will still always be a safe space (facilitated by instructors) that represents only a small taste of work in a mathematical career. How can we provide students with experiences that more rigidly emulate the professional use of mathematics?

The Consortium for Mathematics and its Applications offers two competitive opportunities for students to hone their skills: the Mathematical Contest in Modeling (MCM) and the Interdisciplinary Contest in Modeling (ICM). Over the course of a single weekend, participating teams formulate and communicate a solution to open-ended, broadly-defined problems—such as creating an evacuation plan for the Louvre—for interested stakeholders via a written report. The MathWorks Math Modeling Challenge—sponsored by MathWorks and organized by SIAM—gives high school juniors and seniors 14 hours to tackle a complicated, real-world issue with mathematical modeling and report their results. Participation in such contests enables students to hone their practical mathematical

See **Careers in Industry** *on page 8*

Computational Mechanics

Continued from page 5

flow: nonlinear instability was generally responsible for the weather's unpredictability and fluid turbulence.

But a confounding problem remained. On the one hand, we had time series of measurements of the fluid velocity at a point in a flow. On the other, we had the abstract mathematics of strange attractors — complicated manifolds that circumscribed a system's instability. How could we connect them? Proposals to use the measured time series to "reconstruct" the system's effective state space solved the conundrum. This concept involved extracting the attractor's geometry from a time series. Such reconstruction methods created an effective state space in which to look at the chaotic attractors and quantitatively measure their degrees of instability and attendant complexity. This was finally verified experimentally in 1983, overthrowing the decades-old Landau-Lifshitz multiple incommensurate-oscillator view of turbulence.

Reconstructing a chaotic attractor from a time series became a widely-used technique for identifying and quantifying deterministic chaotic behavior, thus inspiring the field of nonlinear time-series modeling. Yet reconstruction fell short of concisely expressing a system's internal structure. Could we extend reconstruction to extract the system's very equations of motion, enabling us to robustly predict chaotic behavior? A method to reconstruct equations of motion from a data series provided the answer.

This method worked quite well when one happened to choose a mathematical representation that matched the class of nonlinear dynamics generating the behavior. But without the correct representational "basis," it both failed miserably and fell short of revealing how and where to look for a better technique. Thus, even this approach to modeling complex systems was inherently subjective in the choice of representation. Structural complexity remained elusive.

How could one remove this subjectivity? The solution came from a metaphor to the classification scheme for automata developed in discrete computation theory. In the 1950s and 60s, the mathematics of formal languages and automata led to a structural hierarchy of representations that ranged from devices that used finite memory to infinite memories organized in different architectures — tapes, stacks, queues, counters, and the like.

Could we do this for continuous chaotic systems? The answer was framed as a predictive equivalence relation developed from the geometry-of-a-time-series concept of reconstructed state space and adapted to an automata-theoretic setting. The equivalence relation defined a new kind of state, a distribution of futures conditioned on past trajectories in the reconstructed state space. These were the causal states whose resulting probabilistic automata were called ϵ-machines. In this way, many notions of information processing and computing became applicable to nonlinear physics.

Applications

Pithy notation, apt representation, and even the right C++ object class can greatly simplify problem solving. As minimal sufficient statistics, the ϵ-machine causal states play just this role when working with and analyzing complex processes. Figure 1 (on page 5) displays a variety of ϵ-machines, from discrete-time discrete-state to infinite-state and continuous-time processes. Local causal states are adapted to spatiotemporal processes and detect spacetime invariants sets (domains) whose broken symmetries (particles) propagate (see Figure 2).

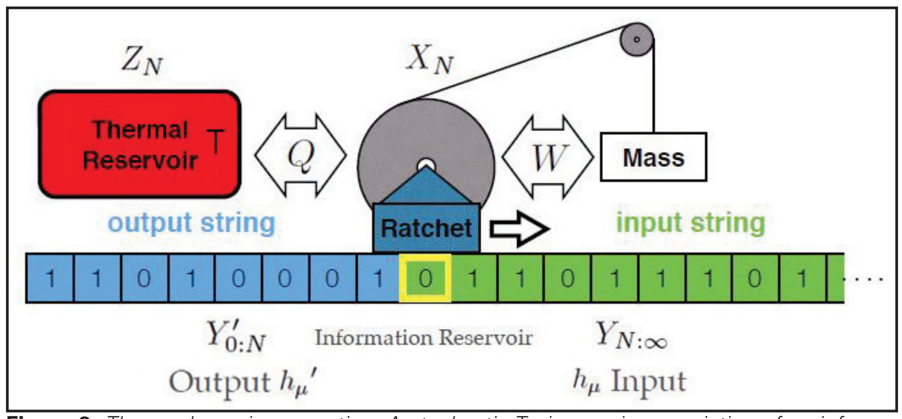


Figure 3. *Thermodynamic computing. A stochastic Turing engine consisting of an information ratchet that processes symbols on a tape using an input-output ϵ-machine. Figure courtesy of Alexander Boyd.*

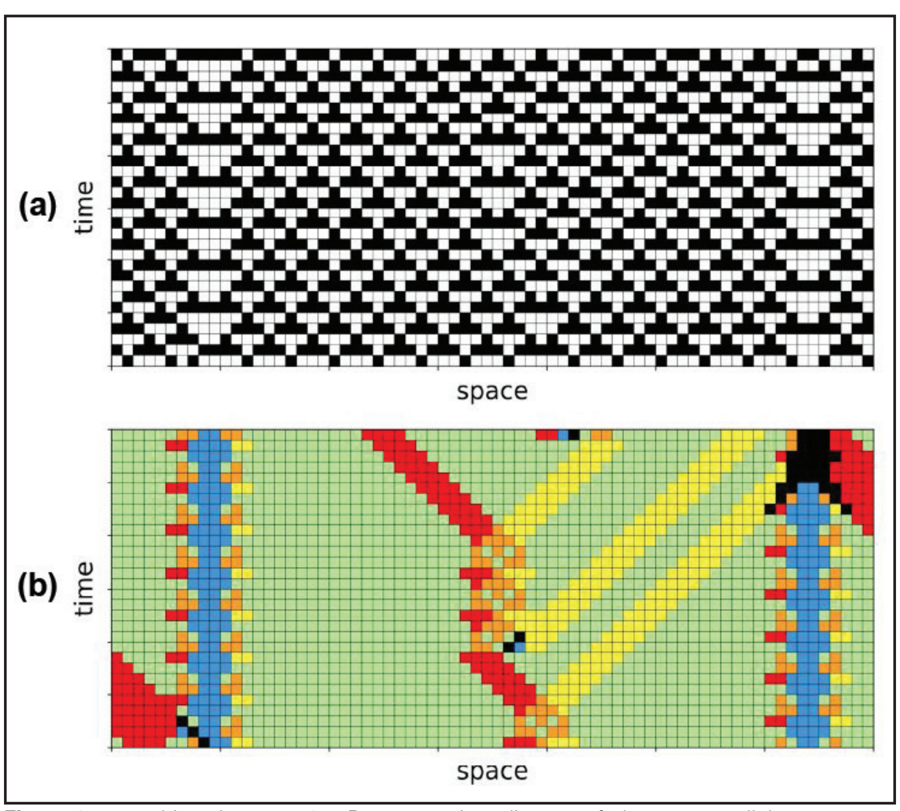


Figure 2. *ϵ-machines in space.* **2a.** *Raw spacetime diagram of elementary cellular automaton 18 evolving from a random initial condition.* **2b.** *Local causal-state spacetime fields calculated from the predictive equivalence relation over past and future lightcones reveal domains and particles. Figure courtesy of Adam Rupe.*

In the mysterious realm of quantum physics, determining a system's quantum causal states allows one to measure quantum structural complexity. Comparison to classical statistical complexity reveals a profound ambiguity: quantum physics and classical physics do not agree on the meaning of structure. That said, quantum ϵ-machines are typically smaller, so future quantum computers will likely provide helpful compression and reduced communication costs.

When combined with recent advances in nanoscale nonequilbrium thermodynamics, ϵ-machines offer the most efficient way to physically implement information processing (see Figure 3). Moreover, the degree of chaos (dynamical instability measured by the Kolmogorov-Sinai entropy rate) controls the rate at which an engine converts disorganized thermal energy into useful work. Additionally, the second law of thermodynamics—when properly extended to account for structure in *information reservoirs*—determines the minimum amount of energy required to perform a given computation.

Conveniently, originally vague intuitive notions about complex systems can now be clearly spelled out. For instance, we now say that a system is *unpredictable* if it has a positive entropy rate: $h_\mu > 0$. A system is considered *complex* if it has positive structural complexity: $C_\mu > 0$. And a system is *emergent* if its structural complexity increases over time: $C_\mu(t') > C_\mu(t)$, if $t' > t$. Finally, we can monitor the amount of internal structure *hidden* from us if its crypticity is positive: $\chi = C_\mu - \mathbf{E} > 0$, where *excess entropy* $\mathbf{E}$ is the shared information between past and future.

Most constructively, and perhaps most telling about causal states, is the manner in which they gave rise to exact methods. Spectral decomposition techniques recently derived from functional calculus yield exact, closed-form expressions for all informational and computational properties of processes generated by ϵ-machines.

Pattern Discovery

Computational mechanics was thus extended and applied far beyond its initial conception 30 years ago. Even so, laying the foundations of a fully-automated "artificial science"—in which theories are automatically built from raw data—remains a challenge. Though the benefits are tantalizing, doing so remains an ambitious goal and its research program is still incomplete. For example, delineating the way in which the causal equivalence relation induces a structural hierarchy of complex processes, rather analogous to the Chomsky hierarchy in discrete symbolic computation theory, is largely unexplored.

That said, the biggest scientific challenges are the recent discoveries that classical and quantum simplicity are ambiguous, and the fact that the most fundamental difficulties arise from our lack of knowledge surrounding statistical dependency, beyond pairwise correlation. In short, Shannon-Kolmogorov information cannot capture all types of statistical dependencies and is mute on many factors of complex system organization. Though these are open problems, I believe that they are also answerable challenges. Perhaps past successes and the style of computational mechanics will inspire their solution.

Acknowledgments: I thank the Santa Fe Institute, the Institute for Advanced Study at the University of Amsterdam, the Telluride Science Research Center, and the California Institute of Technology for their hospitality during visits. This material is based on work supported by, or in part by, Foundational Questions Institute grant FQXi-RFP-1609, the U.S. Army Research Laboratory and the U.S. Army Research Office under contract W911NF-13-1-0390 and grant W911NF-18-1-0028, and via Intel Corporation support of CSC as an Intel Parallel Computing Center.

James P. Crutchfield teaches nonlinear physics at the University of California, Davis, directs its Complexity Sciences Center, and promotes science interventions in nonscientific settings. He is mostly concerned with patterns: what they are, how they are created, and how intelligent agents discover them. His website is http://csc.ucdavis.edu/~chaos/.

Panelists Talk Machine Learning and the Future of Mathematics at ICIAM 2019

By Hans De Sterck, Gitta Kutyniok, James Nagy, and Eitan Tadmor

The excitement and activity surrounding the field of machine learning was clearly evident at the 9th International Congress on Industrial and Applied Mathematics (ICIAM 2019), which took place this summer in Valencia, Spain. Over 25 minisymposia—as well as several prize lectures and invited talks—touched on the theme of "learning," while other invited presentations addressed important mathematical research challenges necessary to advance the field.

A special panel on the future of mathematics in the age of machine learning explored the topic in detail. Panelists Hans De Sterck (University of Waterloo), Gitta Kutyniok (Technische Universität Berlin), James Nagy (Emory University), and Eitan Tadmor (University of Maryland, College Park) represented various core areas of computational and applied mathematics that develop and utilize machine learning techniques, including computational science and engineering, imaging science, linear algebra, and partial differential equations.

Discussion broached a variety of issues surrounding machine learning, such as the obvious fact that machine learning will remain, as mathematician Ali Rahimi stated, "an area comparable to alchemy" without new mathematical understanding and developments. Deep learning is among the most transformative technologies of our time, and its many potential applications—from driverless cars to drug discovery—can have tremendous societal impact. Yet although deep learning retains significant public interest, lay people are largely unaware of the key mathematical and computational challenges in the field. This is particularly true in instances that require many layers for the interpretation of highly complex data patterns. Unfortunately, there are currently no theoretical results to support much of the practical experience suggesting that deep learning algorithms can produce amazing results for high-dimensional data.

Multiple presenters at ICIAM 2019 are attempting to address this gap in mathematical theory by developing novel means of interpreting deep learning as a dynamic optimal control problem with ordinary differential equation and partial differential equation (PDE) models. New mathematical theories will allow concepts from applied mathematics to create a rigorous theoretical basis for designing and training deep neural networks, and subsequently providing insight into their reasoning. A number of minisymposia at the conference focused on the mathematical foundations of deep learning.

Similarly, computational mathematics plays an increasingly important role in machine learning by providing new, efficient optimization algorithms and scalable parallel numerical methods for deep network training. These techniques are essential when training very large networks in ways that scale on high-end parallel computing infrastructure using enormous amounts of data, thus pushing the boundaries of machine learning's capabilities. Several minisymposia at ICIAM 2019 emphasized these novel mathematical developments and their applications in materials, finance, signal and image processing, molecular dynamics, and inverse problems.

The panel generated lively conversation on a multitude of issues, including current limitations of machine learning. For example, more than one attendee expressed concern that machine learning might only be useful "when being wrong is not dangerous." Others noted that limitations revolve primarily around quality of data (i.e., human involvement is typically essential for data labeling, which can cause biases) and the need for massive computing resources. Since current machine learning models are highly susceptible to errors, poor and biased data can induce significant failures. Moreover, the time constraints posed by learning and verification are substantial; because models demand constant training, these computational requirements will only grow with time. Furthermore, if machine learning algorithms are to run on handheld devices (without the need for cloud computing resources, which can have significant latency problems), new approaches that use less data and computing resources are also necessary. Overcoming many of these limitations requires advancements from core areas of computational and applied mathematics, including linear algebra, PDEs, optimization, inverse problems, high-performance computing, statistics, and uncertainty quantification.

At the same time, mathematical disciplines like inverse problems and numerical analysis of PDEs are also impacted by machine learning techniques. Such methods—foremost deep neural networks—can quickly lead to state-of-the-art approaches, particularly for inverse problems in imaging sciences. This occurs due to the dearth of physical models in imaging science, which consequently makes data-driven methods quite effective. From a numerical standpoint, a particularly promising conceptual approach is the use of model-based methods for as long as they are reliable, and the exploitation of learning-type methodologies when they are not. The development of a mathematical underpinning for machine learning and hybrid-type approaches to inverse problems is one main direction of future research. The field of PDEs was slower to embrace machine learning methods, but theoretical results in the high-dimensional regime—typically demonstrating that deep neural networks can overcome the curse of dimensionality—are already available. Research in the aforementioned directions is accelerating.

Students pursuing degrees in areas of data science, including machine learning—whether in mathematics or computer science—must be trained in the relevant mathematical fields. The machine learning revolution and surrounding excitement may slow down in the coming years, but the technology itself is not going to disappear. We can therefore expect the number of undergraduate and graduate students enrolled in computational and applied mathematics courses to significantly increase in the next decade and remain high for the foreseeable future. Moreover, the mathematical community should consider adapting some of its core curricula to include additional topics related to the mathematics of data science. This would be especially beneficial in courses taken by mathematics students, as well as mathematics courses taken by students from other areas of natural and engineering sciences.

Knowledge of machine learning methods is even becoming increasingly important for humanities students. In this sense—and from an educational viewpoint—many ICIAM 2019 attendees are expecting a paradigm shift. Data science and machine learning are rapidly evolving into the leading quantitative and computational endeavors of our time, transforming the way in which society functions. Mathematics, statistics, and mathematics-based algorithms are foundational building blocks of this revolution, and the role and influence of further mathematical developments will only increase as the revolution continues to unfold.

Hans De Sterck is a professor of applied and computational mathematics at the University of Waterloo in Canada. His research focuses on numerical methods for computational science and data science. Gitta Kutyniok is Einstein Professor of Mathematics and head of the Applied Functional Analysis Group at the Technische Universität Berlin. Her research focuses on applied harmonic analysis, compressed sensing, deep learning, imaging science, inverse problems, and numerical analysis of partial differential equations (PDEs). James Nagy is Samuel Candler Dobbs Professor and chair of the Department of Mathematics at Emory University. His research focuses on numerical linear algebra, structured matrix computations, numerical solution of inverse problems, and image processing. Eitan Tadmor is a Distinguished University Professor at the University of Maryland, College Park. His research focuses on theory and computation of PDEs with applications to shock waves, kinetic transport, incompressible flows, image processing, and self-organized collective dynamics.

Careers in Industry

Continued from page 6

skills and obtain a more realistic picture of an applied mathematics career. Furthermore, faculty are beginning to use these programs as templates to design smaller, local competitions. Such contests provide more chances to engage with mathematics, especially for students who might not normally compete in a national or international competition. Those who do participate receive immediate feedback, which is not possible in large-scale competitions like the MCM and ICM.

Finally, industry itself overflows with problems perfect for student engagement. Many institutions collaborate with laboratories, businesses, and industrial companies that allow both undergraduate and graduate students to conduct research in professional settings. Participants gain firsthand experience using mathematics in business and industry, and companies receive help in solving problems while identifying and training potential future hires.

In conclusion, how do we leverage our skills as applied mathematicians to truly train and coach the future generation? We do so by using our unique experiences, knowledge of mathematics' utility, and ability to connect with industry partners and immerse them in meaningful, real-world problems. This is the current trend in applied mathematics education for non-majors, graduate students, and everybody in between. Everyone benefits from focusing on problem-solving, and no group is better prepared to provide contextualized experiences in mathematics than applied mathematicians themselves.

Lucas Castle is the Math that Matters Postdoctoral Fellow in the Applied Mathematics Department at Virginia Military Institute. He can be reached at castlelc@vmi.edu.

Nearly Three Decades at Snowbird: The Iconic Venue and its Influence on Dynamical Systems at SIAM

By Hans Kaper and Marty Golubitsky

After 14 biennial meetings at the Snowbird Ski and Summer Resort just outside of Salt Lake City, Utah, the SIAM Activity Group on Dynamical Systems (SIAG/DS) decided to move its 2021 conference to Portland, Ore. In this article, we reflect on the symbiosis of Snowbird and dynamical systems, and highlight changes that have occurred since the first Snowbird meeting in 1992. Our choice of the term "symbiosis" is deliberate; we believe that the venue—located (almost) at the end of Little Cottonwood Canyon in the Wasatch Range of the Rocky Mountains—has played an important role in the development of an interdisciplinary research community focused on one of the most vibrant areas of applied mathematics.

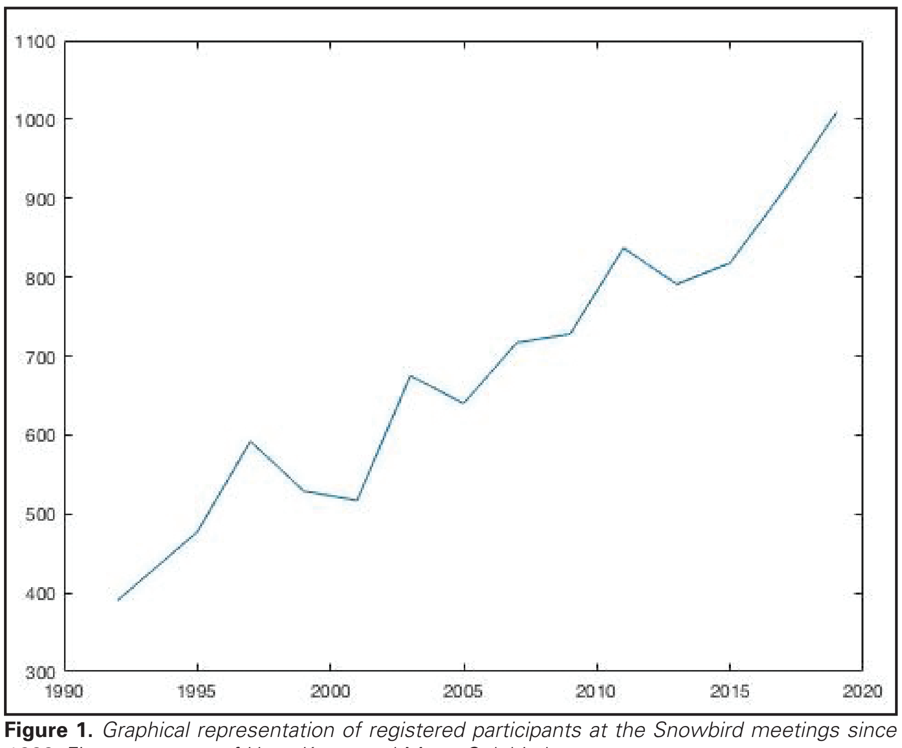


Figure 1. *Graphical representation of registered participants at the Snowbird meetings since 1992. Figure courtesy of Hans Kaper and Marty Golubitsky.*

Why Is Snowbird Different?

SIAG/DS's use of the same venue for more than a dozen consecutive meetings is unique among the SIAGs. All other groups move their meetings among different cities in the U.S.; some even go international. Why is Snowbird different?

A standard caricature of a *Snowbird conference* goes like this: if you want to meet a particular applied dynamicist attending Snowbird, hang out at the SIAM registration desk for 10 minutes. The layout of the meeting rooms on two floors creates a traffic pattern with two critical points where most trajectories intersect; one of those points is also the location of the SIAM desk. Even though the meeting is among the highest attended SIAM conferences, the setup makes it possible to meet people more easily than at most other meetings. Another positive feature of the Snowbird venue is the large number of nooks and crannies where attendees can connect for private scientific discussions. Most conference hotels do not offer this amenity.

Yes, Snowbird is isolated; it is a 45-minute drive from Salt Lake City up the canyon to Cliff Lodge, where the sessions are held. But the scenery is gorgeous, with a breathtaking view of the surrounding mountains from almost anywhere on site. There are essentially no distractions. The meeting focuses on science, and the science is thoroughly interdisciplinary. Other positives include the possibility of skiing (some years) and the ample free time for mingling and networking in the evenings due to the secluded nature of the venue.

Of course, there are negatives. When the number of attendees exceeded the lodge's capacity, late registrants were accommodated at the Alta Ski Area towards the end of the canyon. The alternative was to stay down in the valley, which adds a daily hour-and-a-half commute to and from the conference. Limiting the number of participants would counter SIAM's generally accepted spirit of openness.

Some attendees complained about the lack of evening entertainment at the lodge. At the 2003 SIAM Conference on Applications of Dynamical Systems (DS03), this grievance inspired conference organizers to set up a collection of board games. It is worth noting that we see the dearth of evening activities as both a positive (attendees are readily available) and a negative (it's difficult to separate oneself from them). Other negatives include the altitude (approximately 8,000 feet)—which poses problems for some, especially when combined with jet lag—and the price of food at Snowbird's eateries.

SIAM's established relationship with Snowbird offers further perks. The lodge provides good room rates and many services to its regular customers. The total expense of attending DS21 in Portland will likely not be any cheaper than the cost of attending a Snowbird meeting. The management at Cliff Lodge is familiar with SIAM's requirements, and over the years they have accommodated special requests like the grab-and-go lunch table during midday breaks and the provision of additional meeting rooms.

Oct. 1992: Snowbird (DS92), Peter Bates, Chris K.R.T. Jones
May 1995: Snowbird (DS95), James Meiss, J. D. Crawford
May 1997: Snowbird (DS97), Mary Silber, Steve Strogatz
May 1999: Snowbird (DS99), Emily Stone, Dieter Armbruster
May 2001: Snowbird (DS01), Eric Kostelich
May 2003: Snowbird (DS03), Tasso Kaper, Mary Pugh
May 2005: Snowbird (DS05), Rachel Kuske, Bjorn Sandstede
May 2007: Snowbird (DS07), Sue Ann Campbell, Bernd Krauskopf
May 2009: Snowbird (DS09), Jeff Moehlis, Bruno Eckhardt
May 2011: Snowbird (DS11), Jonathan Dawes, Vivien Kirk
May 2013: Snowbird (DS13), Charles Doering, George Haller
May 2015: Snowbird (DS15), Lora Billings, Panayotis Kevrekidis
May 2017: Snowbird (DS17), Evelyn Sander, Martin Wechselberger
May 2019: Snowbird (DS19), Mason A. Porter, Elaine Spiller

Table 1. *List of every SIAM Conference on Applications of Dynamical Systems that took place at Snowbird, and the chairs/co-chairs for each conference.*

Some Statistics

In May 1990, SIAG/DS held its inaugural meeting in Orlando, Fla., with 427 registered participants. All subsequent meetings took place at Snowbird.

Aside from DS92—the first Snowbird meeting, which took place in October 1992—all SIAG/DS meetings occurred every other May from 1995 until 2019. A list of all meetings, with their co-chairs, is available in Table 1. An additional SIAG/DS-sponsored Pacific Rim Dynamical Systems Conference was held in Maui, Hawaii in August 2000.

The number of registered attendees grew from 390 at DS92 to 1,009 at DS19; the increase was not monotone but certainly substantial (see Figure 1). Beginning with DS13, the number of participants exceeded the guest room capacity of the lodge and late registrants were accommodated elsewhere. The number of meeting rooms at Cliff Lodge also limited the number of minisymposia that the program could feasibly support.

Scientific Programs

The scientific program of a Snowbird meeting typically consisted of invited presentations, minisymposia, contributed papers, and poster sessions. Beginning in 2001, SIAM also awarded the Jürgen Moser Lecture[1] at Snowbird. To accommodate the increasing number of presentations, the length of the meetings grew to five full days with two evening poster sessions.

Table 2 documents the breakdown of presentations over the years. The number of invited presentations decreased to make room for more minisymposia, the number of minisymposia exploded at DS17, and the number of contributed paper sessions diminished in recent years in favor of poster presentations. Parallel sessions were the norm; parallelism increased over the years but was limited by the number of conference rooms at the lodge. Organizing committees had to either allow early and/or late sessions or shorten the length of individual presentations; each committee struggled with this issue and found its own solution. Lodge management helped by enclosing an open area and creating two supplementary breakout rooms.

Conference themes were selected by each organizing committee and varied over the years. It is difficult to verify how strictly the themes were actually followed; they were reflected in the list of invited speakers but functioned more as suggestions for minisymposia organizers and authors of contributed papers. At DS97, the organizing committee made a good-faith effort to classify all minisymposia and contributed paper presentations under six principal themes and ten application areas.[2]

Meeting	IP	MS	CP	PS
DS92	12	33	26	1
DS95	10	46	28	1
DS97	11	51	37	2
DS99	10	55	44	1
DS01	10	66	39	1
DS03	10	110	52	1
DS05	10	110	55	1
DS07	8	96	54	1
DS09	10	124	41	1
DS11	9	143	55	1
DS13	9	136	39	1
DS15	7	134	29	2
DS17	7	179	14	2
DS19	9	186	54	2

Table 2. *Scientific program of Snowbird meetings throughout the years, including invited presentations (IP), minisymposia (MS), contributed papers, (CP), and poster sessions (PS).*

Beginning with DS13 and continuing through DS17, the themes were simply summarized in two sentences: "The scope of this conference encompasses theoretical, computational, and experimental research on dynamical systems. Highlighted areas include the dynamics of biological, chemical, physical, social, and financial systems, along with applications in geophysics, fluid dynamics, engineering, and other applied sciences." DS13 recognized 2013's designation as the year of Mathematics of Planet Earth (MPE), and several presentations addressed MPE-related research topics.

Adaptability has been a striking aspect of Snowbird meetings. In 1995, hot topics included chaos, normal forms, pattern formation, infinite dimensional dynamics, fluid dynamics, and engineering applications. Over two decades later, the 2019 meeting featured more talks on dynamics of networks and applications to biology. This adaptability is due, in part, to the decision to appoint (mostly) junior to mid-career researchers as SIAG/DS program directors, and have both candidates for program director agree to co-chair the Snowbird conference.

Jürgen Moser Lecturers		J.D. Crawford Prize Awardees	
2019	Philip Holmes	2019	Margaret Beck
2017	Edward Ott	2017	Martin Wechselberger
2015	John Guckenheimer	2015	Florin Diacu
2013	Nancy Kopell	2013	Panayotis G. Kevrekidis
2011	James A. Yorke	2011	Eric Vanden-Eijnden
2009	Martin Golubitsky	2009	Arnd Scheel
2007	Harry L. Swinney	2007	Andrew Stuart
2005	Stephen Smale	2005	Dwight Barkley
2003	David P. Ruelle	2003	Yannis G. Kevrekidis
2001	Yakov G. Sinai	2001	Björn Sandstede

Table 3. *Jürgen Moser Lecture prize awardees (left) and J.D. Crawford Prize awardees (right).*

Prizes and Awards

Every two years beginning in 2001, SIAG/DS has awarded two special prizes: the Jürgen Moser Lecture and the J.D. Crawford Prize.[3] As mentioned before, SIAM established the Jürgen Moser Lecture in memory of Jürgen Moser, a leading mathematician who helped develop mathematical theories in celestial mechanics and dynamical systems theory. This prestigious prize is given to a person who has made distinguished contributions to nonlinear

See **Snowbird** *on page 12*

[1] https://www.siam.org/prizes-recognition/activity-group-prizes/detail/full-prize-specifications/siag-ds-jurgen-moser-lecture

[2] https://archive.siam.org/meetings/archives/ds97/overview.htm

[3] https://www.siam.org/prizes-recognition/activity-group-prizes/detail/siag-ds-jd-crawford-prize

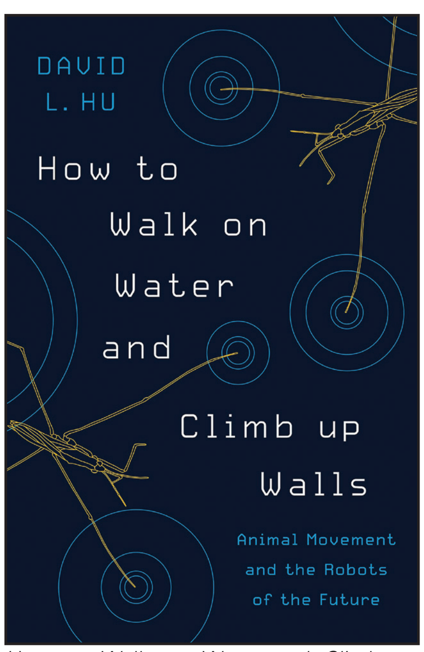
The Physics of Animal Motion

How to Walk on Water and Climb up Walls: Animal Movements and the Robots of the Future. *By David Hu. Princeton University Press, Princeton, NJ, November 2018. 240 pages, $24.95.*

BOOK REVIEW
By Ernest Davis

Animals walk, run, swim, fly, glide, hover, slither, burrow, and swarm. They also regularly move material in and out of their bodies. *How to Walk on Water and Climb up Walls*, a new book by biophysicist David Hu of the Georgia Institute of Technology (Georgia Tech), describes the fascinating science behind the physical and biological principles of animal motion. Hu then links these actions to the engineering of robots that move in similar ways.

The book opens with Jerry, Hu's girlfriend's toy poodle, shaking himself dry after a bath. Jerry shakes about seven times per second and his shaking generates forces up to 12 times gravity; in a fraction of a second, he can eliminate 70 percent of the water in his fur. Hu—with help from his student, Andrew Dickerson—built a wet dog simulator that spun a clipping of Jerry's hair. At both Georgia Tech labs and Zoo Atlanta, the pair took high-speed films of creatures ranging from bears (which shake four times per second) to mice (which shake 29 times per second).



How to Walk on Water and Climb up Walls: Animal Movements and the Robots of the Future. By David Hu. Courtesy of Princeton University Press.

Evolution has managed to design animals that exploit the subtlest properties of the media through which they move. Water striders, as their name suggests, are small insects that walk on water. They are light enough—and their feet are long enough—for the surface tension to support them. If you blow on them gently, they simply glide along the water's surface. Understanding how they move and from where the traction comes was an enigma. Stanford University biophysicist Mark Denny discovered that adult water striders create tiny waves in the water that provide the necessary propulsion. But the legs of infant insects are too small to generate the necessary waves. The mystery of how they move is thus known among water-strider aficionados as "Denny's Paradox." As a project for a fluid mechanics course, Hu determined that another mechanism is involved; the infant water striders' legs produce a tiny vortex in the water, the momentum of which is enough to push the insects forward. Armed with this knowledge, mechanical engineer Brian Chan built a water-strider robot[1] from the aluminum in a soda can. He powered his robot with a thread from an athletic sock, which is also apparently a marvel of modern material design.

Hu proceeds to describe his work—and that of his students, teachers, and collaborators around the world—with creatures of all kinds. He explains how snakes slither across a flat surface, how worms tunnel through mud, and how common sandfish (a type of lizard) burrow through sand. He details the flight mechanism in flying snakes and the altogether different flight apparatus in bumblebees. Hu explores how armies of ants utilize their own bodies to build bridges over gaps, and informs readers of the simple distributed feedback scheme that ants use to decide on bridge placement. He even conducts comparative studies of urination times in different creatures and finds that they are remarkably constant — 10 to 30 seconds in a variety of animals, ranging from his infant son to elephants. Hu and his colleagues investigate the impact of eyelash length in clearing water from the eye (short lashes are much more effective than long lashes), and the role of shark scales in dictating swimming patterns. They analyze the creation of different locomotion mechanisms that require the least energy, and research how—in the right kind of current—a dead fish can "swim" using no energy at all.

These experiments involve both close interaction with all kinds of animals and a wide range of ingenious experimental devices and analytical techniques. Exploring worm motion requires a type of transparent mud, and studying swimming creatures calls for dyed water. Hu and his colleagues utilize high-speed cameras and delicate motion detectors. They employ state-of-the-art computer simulation techniques to calculate the theoretical predictions of fluid and material mechanics, and compare these results to experimental findings. If a scientist can then build a physical robot using the same principles, they would have tangible confirmation of the theory's accuracy (plus a very cool robot). Hu's writing is chatty, entertaining, gracious, and very clear. The book itself is amply illustrated with line drawings and a dozen color plates.

I find it hard to imagine a direction of scientific inquiry with more obvious natural appeal to the general public; Hu's work perfectly fits the "Science is interesting and fun!" campaign that scientists eternally wage. I was therefore shocked—though perhaps in hindsight I should not have been—to learn that Hu's research has been singled out for mockery by the professional obscurantists who infest broadcast news and politics. Why, they cry, should public funds be spent calculating how long it takes elephants to pee? In 2016, *Fox and Friends* featured three of Hu's projects on their "Wheel of Waste." The very fact that this kind of work is so easily understood can become a club to beat it with (most scientific research is protected from this type of attack by the sheer difficulty of explaining what it is about). Even more depressing is the fact that—as Hu discusses towards the end of *How to Walk on Water and Climb up Walls*—scientists working in these areas often struggle to get institutional support at universities or research labs for this type of interdisciplinary work. Perhaps for this reason, Hu's last chapter is somewhat defensive in tone, laying out the potential impact of these studies for both medicine and robotics of all kinds. On the one hand, an enhanced understanding of human motion can lead to better diagnostics (for example, physicians can diagnose Parkinson's disease and similar ailments in part by measuring gait) and improved assistive devices, exercise apparatuses, and prosthetics. On the other hand, the best robot for many tasks is not one that walks like a person, but rather moves in one of the many ways that animals do.

Ernest Davis is a professor of computer science at New York University's Courant Institute of Mathematical Sciences. His book with Gary Marcus, Rebooting AI: Building Artificial Intelligence We Can Trust, was published in September.

[1] In this book, the word "robot" simply means a mechanical toy; there is no requirement that it must be controlled by a computer.

The Cauchy-Schwarz Inequality and a Paradox/Puzzle

MATHEMATICAL CURIOSITIES
By Mark Levi

The Inequality

Consider an asymmetric U-tube — two cylinders connected by a thin tube, as in Figure 1. As I depress the water in the right arm (using a piston, for example), I increase the water's potential energy.[1] Translated into algebra, this becomes the Cauchy-Schwarz inequality, as I will now demonstrate.

The potential energy of a cylindrical column of water of height h and radius r equals the weight times the height $h/2$ of the center of mass, i.e.,

$$kr^2h^2,$$

where $k=\pi\rho g/2$. Here, ρ is the water's density and g is the gravitational acceleration. From here on, we choose units in which $k=1$.

[1] Potential energy increases precisely by the amount of work done to overcome the hydrostatic pressure; friction is of course ignored.

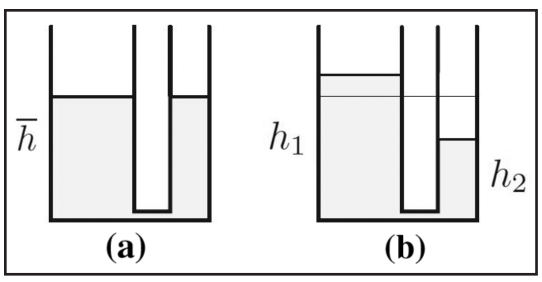


Figure 1. *Potential energy* $P_{(b)} \ge P_{(a)}$; *this is equivalent to (3).*

The new potential energy in Figure 1b is greater:

$$\Sigma r^2h^2 \ge \left(\Sigma r^2\right)\bar{h}^2, \qquad (1)$$

where the subscripts $k=1,2$ are dropped and where $\bar{h}=\Sigma r^2h/\Sigma r^2$ is the average level in Figure 1a. Substituting this value of $\bar{h}$ into (1) gives

$$(\Sigma r^2h^2)(\Sigma r^2) \ge \left(\Sigma r^2h\right)^2. \qquad (2)$$

This is the Cauchy-Schwarz inequality in disguise: setting $rh=x$, $r=y$ implies $r^2h=xy$ and turns (2) into

$$(\Sigma x^2)(\Sigma y^2) \ge (\Sigma xy)^2. \qquad (3)$$

This works verbatim for sums of any number n of terms; one just needs to have n cylinders instead of only two [1].

A Paradox

Consider a symmetric U-tube with water at rest, as in Figure 2. Using a piston, I push the right column of water down; the left column will rise by an equal amount. As much water goes down as up, and by the same distance. Therefore, *the average height of the water—i.e., the height of the center of mass—does not change, and neither does the potential energy.*

But this contradicts the earlier argument, as well as the following thought experiment. Instead of depressing the column, cut a cylinder of water off of the column's top and place this cylinder on top of the other column, thus achieving exactly the same configuration as when using a piston. Since the work done by lifting the cylinder is positive, so is the change in potential energy.

I offer the question in the caption of Figure 2 as a puzzle.

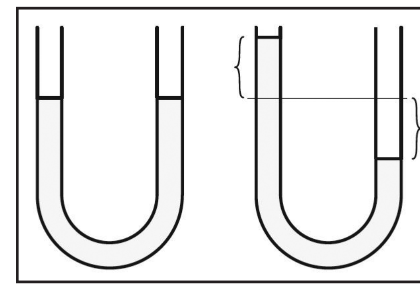


Figure 2. *When I push the water down in one arm of the tube, just as much water moves up the other arm, and by the same distance. In other words, the average vertical displacement is zero. The height of the center of mass thus remains unchanged. Where is the mistake in this short argument?*

The figures in this article were provided by the author.

References

[1] Levi, M. (2019). A water-based proof of the Cauchy-Schwarz inequality. *Am. Math. Monthly*, in press.

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

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" at the SIAM website (**www.siam.org**) or proceed directly to **www.siam.org/careers**.*

California Institute of Technology

Computing and Mathematical Sciences Department

The Computing and Mathematical Sciences (CMS) Department at the California Institute of Technology (Caltech) invites applications for tenure-track faculty positions in all areas of applied mathematics, computer science, and related disciplines.

Areas of interest include (but are not limited to) algorithms, data assimilation and inverse problems, dynamical systems and control, geometry, machine learning, mathematics of data science, networks and graphs, numerical linear algebra, optimization, partial differential equations, probability, scientific computing, statistics, stochastic modeling, and uncertainty quantification. Application foci include computational physical and life sciences, distributed systems, economics, graphics, quantum computing, and robotics and autonomous systems. The CMS Department is part of the Division of Engineering and Applied Science (EAS), comprising researchers working in and between the fields of aerospace, civil, electrical, environmental, mechanical, and medical engineering, as well as materials science and applied physics. The institute as a whole represents the full range of research in biology, chemistry, engineering, physics, and the social sciences.

A commitment to world-class research, as well as high-quality teaching and mentoring, is expected, and appointment as an assistant professor is contingent upon the completion of a Ph.D. degree in applied mathematics, computer science, engineering, or the sciences. The initial appointment at the assistant professor level is four years and is contingent upon completion of a Ph.D. degree. Reappointment beyond the initial term is contingent upon successful review conducted prior to the commencement of the fourth year.

Applications will be reviewed beginning November 15, 2019, but all applications received before January 15, 2020 will receive full consideration. For a list of required documents and full instructions on how to apply online, please visit **https://applications.caltech.edu/jobs/cms**. Questions about the application process may be directed to **search@cms.caltech.edu**.

Caltech is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability status, protected veteran status, or any other characteristic protected by law.

Eleanor and Howard Morgan Professorship
School of Operations and Information Engineering

The School of Operations Research and Information Engineering at Cornell University invites exceptional applicants for the Eleanor and Howard Morgan Professorship. The School of ORIE seeks visionary candidates who will drive new research thrusts, catalyze growth in existing areas, and help lead efforts to realize the School's vision of a world in which operation research and analytics are fundamental to improved decision-making that helps address some of society's most pressing problems.

Cornell ORIE is a diverse group of high-quality researchers and educators interested in probability, optimization, statistics, machine learning, simulation, and a wide array of applications including, but not limited to, e-commerce, supply chains, scheduling, manufacturing, transportation and mobility, health care, financial engineering, service systems and network science. We value mathematical and technical depth and innovation, and experience with applications and practice. Ideal candidates will have correspondingly broad training and interests. The School of Operations Research and Information Engineering spans both the traditional Cornell campus in Ithaca and the new presence at Cornell Tech on Roosevelt Island in New York City. This position is for the Cornell Ithaca campus, but the potential for connection with the Cornell Tech campus and its associated mission of external engagement is desirable.

Applicants should submit a curriculum vitae, statement of research and leadership vision, and complete contact information for at least three references. Applications will be accepted until the position is filled. Submit applications online at **https://academicjobsonline.org/ajo/jobs/14253**. Questions may be addressed by email to David Shmoys, Chair, Eleanor and Howard Morgan Chair Search Committee at **morganchairsearch@cornell.edu**.

ORIE and the College of Engineering at Cornell embrace diversity and seek candidates who can contribute to a welcoming climate for students, staff and faculty of all races and genders. Cornell University seeks to meet the needs of dual career couples, has a Dual Career program, and is a member of the Upstate New York Higher Education Recruitment Consortium to assist with dual career searches. Visit **www.unyherc.org/home** to see positions available in higher education in the upstate New York area.

Cornell University is an innovative Ivy League university and a great place to work. Our inclusive community of scholars, students and staff impart an uncommon sense of larger purpose and contribute creative ideas to further the university's mission of teaching, discovery and engagement. With our main campus located in Ithaca, NY Cornell's far-flung global presence includes the medical college's campuses in Manhattan and Doha, Qatar, as well as the new Cornell Tech campus located on Roosevelt Island in the heart of New York City.

Diversity and Inclusion are a part of Cornell University's heritage. We are a recognized employer and educator valuing AA/EEO, Protected Veterans, and Individuals with Disabilities. We strongly encourage women and underrepresented minorities to apply.

Low Precision Floating-Point Formats: The Wild West of Computer Arithmetic

By Srikara Pranesh

Floating-point arithmetic is fundamental to scientific computing and lies at the heart of almost all numerical computations. The 1985 IEEE standard 754 for floating-point arithmetic marked the end of a turbulent period in scientific computing, during which vendors had their own implementations of floating-point arithmetic. All hardware vendors gradually adopted the IEEE standard of single and double precisions.

Until recently, the landscape of floating-point arithmetic—following employment of the IEEE standard—largely remained the same. However, hardware continually advanced to achieve higher performance, and numerical libraries evolved to efficiently use the hardware. Jack Dongarra and his colleagues [4] demonstrated this progression for singular value decomposition and deduced that *communication is far more expensive than computation*. Therefore, algorithms that minimize communication at the expense of increased computation are the norm; numerical libraries like PLASMA [5] are based on this philosophy. One thing that remained constant during all of these developments was the floating-point formats, and this was about to change.

The 2008 revision of the IEEE 754 standard introduced half precision (or fp16) as a storage format. This was meant to reduce the cost of data movement, as it is cheaper to move 16 bits of data than 32 or 64 bits. However, once half precision was deemed sufficient for deep learning applications, researchers began using fp16 for computation, with a natural extension of the arithmetic rules. Half precision is now available on the NVIDIA P100 (2016) and V100 (2017) graphics processing units (GPUs), as well as the AMD Radeon Instinct MI25 GPU (2017). Although fp16 offers massive speedups, the maximum value it can represent is approximately 65,500, thus making overflow very likely. To address this issue, Google proposed an alternative half-precision format called the bfloat16.[1] Properties of fp16 and bfloat16 are displayed in Table 1.

The range of bfloat16—the format currently used in Google tensor processing units (TPUs)—is similar to single precision but has a lower precision than fp16. Intel will support bfloat16 in its upcoming Nervana Neural Network Processor and Cooper Lake processors. To further accelerate deep learning applications, an eight-bit floating-point format is also under consideration [12]. Additionally, researchers are contemplating nonstandard rounding modes—such as stochastic rounding—to enhance computational accuracy with these low-precision formats [9]. Another interesting technological innovation is the block-fused multiply-add unit, which can perform

$$C + A \times B, \quad A, B, C \in \mathbb{R}^{n \times n} \qquad (1)$$

in a single clock cycle with one rounding error for some specific value of n. This feature is already available in the tensor cores of NVIDIA V100 (where $n = 4$). The Summit machine at Oak Ridge National Laboratory (ORNL), which leads the latest Top500 lists,[2] comprises 27,000 V100s and has achieved an exaop performance using the tensor cores. Furthermore, 133 systems in the June 2019 Top 500 list employ accelerators, over 73 percent of which use GPUs that support fp16.[3] Multiprecision computing units called matrix units (MXU), which operate on 128×128 matrices, are present in Google TPUs as well. However, Google TPUs are not commercially accessible, and details of MXU computation are not publicly available.

With regard to future machines, the Japanese Fugaku exascale machine will be based on the A64FX ARM processor with fp16 support. The Frontier exascale machine—to be installed at ORNL—will use AMD GPUs, which support fp16.[4] In short, GPUs and low-precision formats are here to stay and have transformed the friendly neighbourhood of floating-point arithmetic into the Wild West. Development of algorithms that can exploit these new floating-point formats is therefore of great interest.

	u	$x^s_{\min}$	$x_{\min}$	$x_{\max}$
bfloat16	3.91×10^{-3}	9.18×10^{-41}	1.18×10^{-38}	3.39×10^{38}
fp16	4.88×10^{-4}	5.96×10^{-8}	6.10×10^{-5}	6.55×10^{4}

Table 1. *Parameters for bfloat16, fp16 arithmetic, to three significant figures: unit roundoff u, smallest positive (subnormal) number $x^s_{\min}$, smallest normalized positive number $x_{\min}$, and largest finite number $x_{\max}$. Intel's bfloat16 specification does not support subnormal numbers.*

In the field of numerical linear algebra, Erin Carson and Nicholas Higham have proposed an algorithm for the solution of a linear system of equations that is given in double precision using fp16 [2]. They perform lower-upper (LU) factorization in fp16 and solve the update equation of iterative refinement via the generalized minimal residual method (GMRES), with the low-precision LU factors as preconditioners. A speedup of up to four over-highly-optimised libraries using the tensor cores of NVIDIA V100 has been demonstrated [6]. The algorithm achieved a performance of 445 petaflops—almost three times that of an optimised double-precision solver—when solving a dense linear system of 10 million equations at scale on the Summit machine.

Several matrices appearing in actual applications have entries that exceed the overflow limit of fp16. For example, many metals' modulus of elasticity is $\mathcal{O}(10^9)$. To address this issue, researchers have proposed a scaling algorithm with application to the solution of a linear system [8]. Even with the enormous computing power already available, it is still impossible to run very high-fidelity simulation models in climate studies, which can predict the extent of the effects of global warming [10]. Therefore, scientists are contemplating multiprecision ideas to solve climate models of higher fidelity [3]. To enhance the speed of Monte Carlo simulations, researchers are considering representing the samples in low precision, with applications in Ising models [13] and finance [1].

Higham wrote about the challenges and potential benefits of multiprecision algorithms in a previous issue of *SIAM News* [7]. The two years since his article have seen further changes in the landscape of floating-point arithmetic because of architectural advancements like tensor cores. In 2005, Herb Sutter announced the advent of multicore architectures and proclaimed that "the free lunch is over" [11]. The onset of hardware that supports low precision marks the end of yet another free lunch, as new algorithms—rather than software optimisation—are the key to extracting benefits from such hardware.

References

[1] Belletti, F., King, D., Yang, K., Nelet, R., Shafi, Y., Chen, Y., & Anderson, J. (2019). Tensor processing units for financial Monte Carlo. Preprint, *arXiv:1906.02818*.

[2] Carson, E., & Higham, N.J. (2018). Accelerating the solution of linear systems by iterative refinement in three precisions. *SIAM J. Sci. Comput., 40*(2), A817-A847.

[3] Dawson, A., Düben, P.D., MacLeod, D.A., & Palmer, T.N. (2018). Reliable low precision simulations in land surface models. *Clim. Dyn., 51*(7), 2657-2666.

[4] Dongarra, J., Gates, M., Haidar, A., Kurzak, J., Luszczek, P., Tomov, S., & Yamazaki, I. (2018). The singular value decomposition: Anatomy of optimizing an algorithm for extreme scale. *SIAM Rev., 60*(4), 808-865.

[5] Dongarra, J., Gates, M,. Haidar, A., Kurzak, J., Luszczek, P., Wu, P.,…, Relton, S. (2019). PLASMA: Parallel linear algebra software for multicore using OpenMP. *ACM Trans. Math. Soft., 45*(2), 16-35.

[6] Haidar, A., Tomov, S., Dongarra, J., & Higham, N.J. (2018). Harnessing GPU tensor cores for fast FP16 arithmetic to speed up mixed-precision iterative refinement solvers. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis* (pp. 47:1-47:11). Dallas, TX: IEEE Press.

[7] Higham, N.J. (2017). A multiprecision world. *SIAM News, 50*(8), p. 2.

[8] Higham, N.J., Pranesh, S., & Zounon, M. (2018). Squeezing a matrix into half precision, with an application to solving linear systems. *SIAM J. Sci. Comp., 41*(4), A2536-A2551.

[9] Hopkins, M., Mikaitis, M., Lester, D.R., & Furber, S. (2019). Stochastic rounding and reduced-precision fixed-point arithmetic for solving neural ODEs. Preprint, *arXiv:1904.11263*.

[10] Palmer, T. (2014). Climate forecasting: Build high-resolution global climate models. *Nat. News, 515*(7527), 338.

[11] Sutter, H. (2005). The free lunch is over: A fundamental turn toward concurrency in software. *Dr. Dobb's Journ., 30*(3), 202-210.

[12] Wang, N., Choi, J., Brand, D., Chen, C.-Y., & Gopalakrishnan, K. (2018). Training deep neural networks with 8-bit floating point numbers. *Adv. Neur. Info. Process. Syst., 31*, 7686-7695.

[13] Yang, K., Chen, Y.-F., Roumpos, G., Colby, C., & Anderson, J. (2019). High performance Monte Carlo simulation of Ising model on TPU clusters. Preprint, *arXiv:1903.11714*.

Srikara Pranesh is a research associate in the Department of Mathematics at the University of Manchester. His research interests mainly include numerical analysis and numerical linear algebra.

[1] https://en.wikipedia.org/wiki/Bfloat16_floating-point_format
[2] https://www.top500.org
[3] https://www.top500.org/statistics/list/
[4] https://www.amd.com/en/graphics/servers-radeon-instinct-mi

Snowbird

Continued from page 9

science. The awardee delivers a plenary lecture at the biennial SIAG/DS conference.

The J.D. Crawford Prize was created in honor of John David Crawford, a co-organizer of DS95, for his contributions to the physics of collisionless plasmas and pattern formation. The prize recognizes one individual for his/her recent outstanding work on a topic in nonlinear science. Table 3 (on page 9) lists recipients of both prizes.

In addition, SIAG/DS presents the Red Sock Award[4] for the best poster presentations by students or postdoctoral researchers at Snowbird. Four awards of equal merit are made at the end of each meeting. The prize honors James A. Yorke, and each winner receives a pair of red socks.

Outlook

It takes a while to get to know the positives and negatives of any place, so we'll keep our fingers crossed for DS21. We hope that the new Portland venue will provide some of the intimacy and attendee accessibility that Snowbird has delivered for the past 27 years. Those of us who have attended multiple Snowbird meetings will remain grateful for having had that opportunity.

Hans Kaper, founding chair of the SIAM Activity Group on Mathematics of Planet Earth and editor-in-chief of SIAM News, is affiliate faculty in the Department of Mathematics and Statistics at Georgetown University. He is a former chair of the SIAM Activity Group on Dynamical Systems (SIAG/DS). Marty Golubitsky is a distinguished professor of mathematics at the Ohio State University. He is a past president of SIAM, the founding editor-in-chief of the SIAM Journal on Applied Dynamical Systems, and a former chair of SIAG/DS.

[4] https://www.siam.org/prizes-recognition/activity-group-prizes/detail/siag-ds-red-sock-award

The Cliff Lodge at the Snowbird Ski and Summer Resort in the Wasatch Range of the Rocky Mountains outside of Salt Lake City, Utah. For the past 27 years, the SIAM Conference on Applications of Dynamical Systems took place at this venue. Photo courtesy of Hans Kaper.