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

siam news

Volume 56/ Issue 10
December 2023

Computational Attorneys: Opportunities and Challenges

By Frank Schilder and Dell Zhang

The idea of computers that serve as attorneys once seemed like a distant, futuristic concept. However, the advent of large language models means that this seemingly far-off notion might actually be on the horizon. For many years, the legal domain has been an important application field for various data mining and machine learning technologies. While some progress has occurred in the automation of legal compliance management and legal information processing, most advances have focused on specific, well-defined tasks (e.g., the classification or clustering of legal text, legal entity extraction, and legal information retrieval).

We coin the term *computational attorney* as an intelligent software agent that embodies the future capabilities of these systems and can assist human lawyers with a wide range of complex, high-level tasks, such as drafting legal briefs for prosecution or defense. Here, we reexamine the current research agenda on legal artificial intelligence (AI) by asking a critical question: What does it take to make a computational attorney?

Past and Present

Research on legal AI began in the 1990s when scientists started to utilize data mining and machine learning to address basic legal tasks, like the computer-assisted classification of legal abstracts. Over the years, researchers have harnessed these techniques to automate other minor and repetitive legal projects, including similar case matching, litigation analytics, and information extraction from legal documents [1, 5, 9, 10].

The next big step in the evolution of legal AI was the creation of foundation models for a variety of legal tasks; LEGAL-BERT is one such example [3]. In our opinion, the pinnacle of this line of research is a *Large Legal Language Model* (L^3M) that is *pretrained* on an extensive legal corpus. The collection of a massive amount of quality text data in the legal domain, which helps to train and tune such a model, is key to its success.

In addition to handling extremely high labeling costs for demanding legal tasks via zero-shot or in-context learning, L^3Ms can also accommodate possible ambiguities and idiosyncrasies to meet the challenges of thoroughness and specialized knowledge in the legal domain. The continued maturation of L^3Ms will achieve new heights in legal natural language processing. When an L^3M reaches a certain scale, for example, it may start to exhibit "emergent abilities" — one of which is *legal reasoning* [11]. L^3M-based *legal prompt engineering* has demonstrated impressive performance and potential on legal text entailment and can even answer questions from bar exams [6, 12].

Looking to the Future

While ongoing advancements in L^3Ms are encouraging, several challenges currently preclude them from becoming fully functional computational attorneys. We present these challenges as actual measurements for the required performance of a system that uses an L^3M. However, it is important to note that future research may identify barriers that today's transformer-based models cannot overcome.

First, these models must be *updatable* to keep pace with changes in the legal domain. Timely updates with novel information can significantly enhance a model's ability to deliver reliable solutions. New methods to revise a model in the post-deployment stage are preferable to retraining it from scratch, which is both expensive and time-consuming. Continual learning, machine unlearning, and data stream mining are examples of techniques that can facilitate this process.

Second, these models must be *stable*. They should be able to reason within the

See **Computational Attorneys** *on page 3*

Figure 1. *Artificial intelligence meets legal expertise: a computational attorney at work. Figure courtesy of iStock.com/style-photography.*

AI for Science

By Weinan E

Scientific research serves two major purposes: (i) Discovering fundamental principles, such as the laws of planetary motion and the laws of quantum mechanics, and (ii) solving practical problems, like those that arise in engineering and industry. Researchers typically utilize two major approaches: the Keplerian paradigm (the data-driven approach) and the Newtonian paradigm (the first-principles-driven approach). Johannes Kepler's discovery of the laws of planetary motion is the best example of the former. Kepler found these laws by analyzing experimental data; later, using the laws of mechanics and gravitation, Isaac Newton was able to reduce the problem of planetary motion to ordinary differential equations (ODEs) and derive Kepler's laws. In short, Kepler first made the discovery but did not understand the reasons behind it, so Newton took it one step further and discovered the fundamental principles — which are applicable to many other problems.

For practical purposes, the task of finding first principles was basically accomplished with the establishment of quantum mechanics. In 1929, Paul Dirac declared that "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble" [1]. His comment applies not just to chemistry but also to biology, materials science, and numerous other scientific and engineering disciplines. In practical situations, we can often use simplified principles—such as Euler's equations for gas dynamics and the Navier-Stokes equations for fluid dynamics—instead of relying on quantum mechanics.

With these fundamental principles in hand, essentially all natural science and related engineering problems reduce to mathematical problems — specifically to ODEs or partial differential equations (PDEs). Unfortunately, before the development of effective tools, the only thing that scientists can do to solve these practical problems is to simplify or ignore the principles.

The first major advance in this realm occurred when John von Neumann recognized that computers and numerical algorithms should allow us to directly utilize these fundamental principles in a practical way. Researchers have since developed many numerical algorithms—such as the finite difference method, finite element method, and spectral methods—to solve the corresponding PDEs. The basic starting point of these algorithms is the fact that general functions can be approximated by polynomials or piecewise polynomials. The corresponding impact has been tremendous. Scientific computing has become the foundation of modern technology and engineering science; in fact, the introduction of numerical algorithms has revolutionized countless disciplines, from structural mechanics and fluid mechanics to electromagnetism.

However, many problems cannot yet be treated in this way. For instance, we are still quite far from employing first principles to successfully address material properties, materials design, and drug design. In these types of areas, theoretical work is usually fairly detached from the real world; real-world problems must instead be solved empirically (by trial and error).

All of these "hard" problems share one common factor: they depend on many variables and thus suffer from the *curse of dimensionality*. For example, consider Schrodinger's equation in quantum mechanics. Neglecting symmetry, the number of independent variables for the wave function is three times the number of particles. A system with 10 electrons is extremely simple, but a PDE in 30 dimensions is highly nontrivial.

It is here that deep learning might be able to help. Deep learning successfully classifies images, generates fake pictures of human faces, and produces Go strategies that defeat the best human players. While

See **AI for Science** *on page 2*

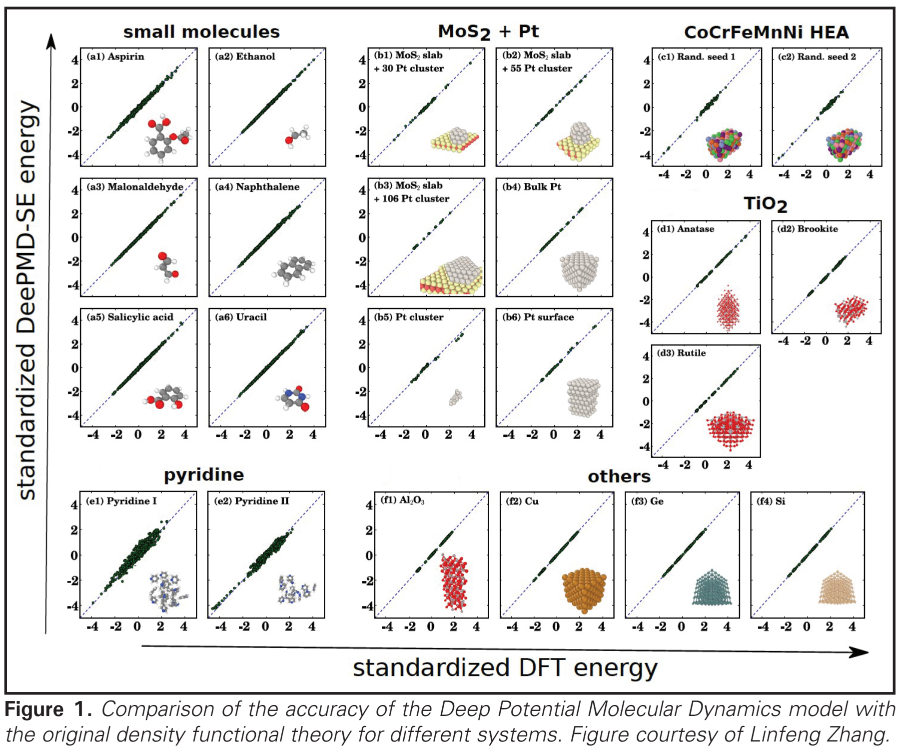


Figure 1. *Comparison of the accuracy of the Deep Potential Molecular Dynamics model with the original density functional theory for different systems. Figure courtesy of Linfeng Zhang.*

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 56/ Issue 10/ December 2023

AI for Science

Continued from page 1

these scenarios are common applications of artificial intelligence (AI), they respectively focus on approximating functions, approximating and sampling probability distributions, and solving Bellman's equations — all of which are standard problems in applied mathematics. However, these AI problems have much higher dimensions than standard applied math problems.

Deep learning's effective performance on these problems suggests that deep neural networks are particularly successful at approximating functions in high dimensions. While a complete mathematical theory for deep learning does not yet exist, we do have some hints as to why this is the case. If we approximate a function with piecewise linear functions on a regular mesh, the error is proportional to the square of the mesh size; this is the origin of the curse of dimensionality. But if we instead approximate the function with neural network functions, we can show—at least in some situations—that the error rate does not deteriorate with dimension [4].

This observation has several important implications. Because functions are among the most basic mathematical objects, a new tool that approximates them in high dimension will impact many different areas. In particular, deep learning can help solve the aforementioned problems that suffer from the curse of dimensionality. This is the starting point of *AI for science*.

The most successful example in this direction is AlphaFold:[1] an AI program from Google DeepMind.[2] By exploiting protein sequence datasets and the most advanced deep learning models, DeepMind shocked the world by developing the AlphaFold 2 algorithm that elegantly solved the protein structure problem [6].

Although AlphaFold 2 is solely driven by data, AI for science is not a purely data-driven paradigm. In fact, the main difference between AI and science is the presence of a set of first principles in science (as discussed earlier). One major component in AI for science is the use of AI to develop better algorithms or approximate models for these first principles. In this regard, the best-known example occurs in molecular dynamics — a fundamental tool in biology, materials science, and chemistry. The idea is to study the properties of molecules and materials by examining the dynamic trajectories of the atoms in the system (the dynamics of the atoms simply follow Newton's law). The difficult part involves modeling the forces between the atoms, which are governed by the *interatomic potential*. In the past, scientists have either tried to guess an approximate functional form of the interatomic potential (the empirical approach) or used quantum mechanics models to compute the forces on the fly (the *ab initio* approach). The former is unreliable and the latter is quite expensive.

Machine learning offers a new paradigm that only uses quantum mechanics to supply the data. Based on that data, we can employ machine learning to create an accurate approximation of the interatomic potential, then utilize that approximation to perform molecular dynamics simulations.

To truly make this technique work, we must address two important issues. The first issue pertains to network structure, which should be extensive and respect physics. *Being extensive* allows us to perform learning on small systems and use the results for larger systems; *respecting physics* means that we need to keep the symmetries, conservation laws, and other physical constraints in place. In the current context, the main problem involves maintaining the translational, rotational, and permutational symmetries [8].

The second issue concerns data. If we ultimately seek an approximate potential energy function that performs as well as the original quantum mechanics model in all practical situations, the training dataset must be able to represent these situations. However, we want the dataset to be as small as possible since the quantum mechanics computations that calculate the data are expensive. This conundrum calls for an adaptive data generation scheme—such as the exploration-labeling-training algorithm—that generates data on the fly as learning takes place [9].

Given these requirements, we can indeed produce neural network approximations of the interatomic potential with *ab initio* accuracy for a large class of (if not all) atomic systems. One relevant model is called Deep Potential Molecular Dynamics (DeePMD) (see Figure 1, on page 1). Together with high-performance computing, DeePMD has extended our ability to perform molecular dynamics with *ab initio* accuracy on systems with thousands of atoms to systems with billions of atoms [3, 5]. The DeePMD software package DeePMD-kit[3] facilitates the use of DeePMD with minimal effort [7].

We can apply similar ideas to other modeling schemes. For example, we can utilize highly accurate quantum chemistry computational data to train density functional theory models that are more universal and accurate. We can also develop more accurate and reliable coarse-grained molecular dynamics models, more accurate moment closure models for kinetic equations, and so forth. In fact, machine learning serves as the missing tool for multiscale, multiphysics modeling.

AI techniques can also enhance our experimental capabilities by providing inversion algorithms that are more efficient and accurate. AI-based algorithms deliver realistic and accurate data for the forward problem, and we can exploit the differentiable structure in AI-based formulations to solve the inverse problem. This line of work is still at a preliminary stage, but it will undoubtedly change the way in which experiments are conducted and experimental apparatuses are designed.

In addition, AI will likely have a serious impact on the handling of literature and other scientific knowledge. These sources are major suppliers of inspiration for our research, but finding and studying them is a highly nontrivial process. We can thus imagine the use of AI—such as AI databases and large language models—to collect and query the existing literature in a more efficient manner.

With these possibilities comes a new "Android paradigm" for scientific research. In this novel paradigm, the scientific community will work together to build an infrastructure that includes AI-based algorithms for physical models, AI-enhanced experimental facilities, and an innovative knowledge database that encompasses the literature and other scientific data. These platforms constitute the "Android platform" for scientific research. Scientists can then organize work on specific applications—like the search for catalysts in a particular reaction or the design of new batteries—on top of this Android platform. This horizontally integrated approach should undoubtedly accelerate the process of scientific research, help break disciplinary barriers, and enhance interdisciplinary research and education.

With this vision in mind, we initiated the DeepModeling open source platform[4] in 2018. This platform invites the scientific community to work together and build the AI-enhanced infrastructure for physical modeling and data analysis. It has since attracted hundreds of developers and facilitated more than 40 projects under active collaboration.

Such a paradigm shift will also bring fundamental changes to the field of applied mathematics. After all, applied math has always centered on the development of tools for first-principles-driven and data-driven approaches [2]. The necessary ingredients for the Android platform are also the major components of applied mathematics. As science becomes increasingly integrated under this platform, applied math will likely become the foundation of interdisciplinary research.

References

[1] Dirac, P.A.M. (1929). Quantum mechanics of many-electron systems. *Proc. R. Soc. Lond. A*, *123*(792), 714-733.

[2] E, W. (2021). The dawning of a new era in applied mathematics. *Not. Am. Math. Soc.*, *68*(4), 565-571.

[3] E, W., Han, J., & Zhang, L. (2021). Machine learning-assisted modeling. *Phys. Today*, *74*(7), 36-41.

[4] E, W., Ma, C., Wu, L., & Wojtowytsch, S. (2020). Towards a mathematical understanding of neural network-based machine learning: What we know and what we don't. *CSIAM Trans. Appl. Math.*, *1*, 561-615.

[5] Jia, W., Wang, H., Chen, M., Lu, D., Lin, L., E, W., … Zhang, L. (2020). Pushing the limit of molecular dynamics with *ab initio* accuracy to 100 million atoms with machine learning. In *SC '20: Proceedings of the international conference for high performance computing, networking, storage and analysis*. IEEE Press.

[6] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*, 583-589.

[7] Zeng, J., Zhang, D., Lu, D., Mo, P., Li, Z., Chen, Y., … Wang, H. (2023). DeePMD-kit v2: A software package for deep potential models. *J. Chem. Phys.*, *159*(5), 054801.

[8] Zhang, L. Han, J., Wang, H., Saidi, W., Car, R., & E, W. (2018). End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. In *Advances in neural information processing systems 31 (NeurIPS 2018)*. Montreal, Canada. Curran Associates, Inc.

[9] Zhang, L., Wang, H., & E, W. (2018). Reinforced dynamics for enhanced sampling in large atomic and molecular systems. *J. Chem. Phys.*, *148*(12), 124113.

Weinan E is director of the AI for Science Institute in Beijing and a professor in the Center for Machine Learning Research and the School of Mathematical Sciences at Peking University. His main research interests are numerical algorithms, machine learning, and multiscale modeling with applications in chemistry, materials science, biology, and fluid mechanics.

[1] https://alphafold.ebi.ac.uk
[2] https://deepmind.google
[3] https://github.com/deepmodeling/deepmd-kit
[4] https://github.com/deepmodeling

siam news

ISSN 1557–9573. Copyright 2023, all rights reserved, by the Society for Industrial and Applied Mathematics, SIAM, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688; (215) 382-9800; siam@siam.org. To be published 10 times in 2023: January/February, March, April, May, June, July/August, September, October, November, and December. The material published herein is not endorsed by SIAM, nor is it intended to reflect SIAM's opinion. The editors reserve the right to select and edit all material submitted for publication.

Advertisers: For display advertising rates and information, contact the Department of Marketing & Communications at marketing@siam.org.

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

An Internship Network in the Mathematical Sciences

By Richard Laugesen and Fadil Santosa

Most mathematical sciences departments—especially at research universities—primarily prepare Ph.D. students for academic careers. However, the substantial growth in Ph.D. programs across the U.S. means that the number of Ph.D. students who graduate each year far exceeds the number of available tenure-track placements. After a few years in postdoctoral positions, many students belatedly realize that careers in business, industry, and government (BIG) offer a myriad of promising opportunities.

This mismatch between academic preparation and graduates' career pathways has troubled us for some time. In 2013, Richard Laugesen, Yuliy Baryshnikov, and Lee DeVille of the University of Illinois Urbana-Champaign (UIUC) created the Program for Interdisciplinary and Industrial Internships at Illinois[1] (PI4) to provide technical training, open-ended summer research opportunities, and a limited number of interdisciplinary or industrial internships at local companies. The PI4 program was funded by the National Science Foundation (NSF) and had a major impact on participating students. Two years later and in pursuit of similar goals, Fadil Santosa and Daniel Spirn (then both of the University of Minnesota) created the Math-to-Industry Boot Camp[2] within the Institute for Mathematics and its Applications; this program continues to draw participants from around the country who receive technical training and take part in industry-sponsored group projects.

With encouragement from Philippe Tondeur, we built upon these initiatives to develop a concept that is loosely modeled on Canada's wildly successful Mitacs program.[3] We pitched our idea to NSF's Division of Mathematical Sciences, and their positive feedback prompted us to submit an unsolicited proposal to their Research Infrastructure Program.[4] The proposal sought to prepare students from a variety of universities for internships that would result from our intensive efforts to establish relationships with companies. We received funding for the Internship Network in the Mathematical Sciences[5] (INMAS) in 2020, and the networks' two branches—centered at UIUC and Johns Hopkins University (Santosa's current institution)—soon sprang into action.

About INMAS

INMAS launched in the early days of the COVID-19 pandemic. The Midwestern branch initially consisted of the hub at UIUC and nodes at the University of Illinois Chicago, Indiana University, and Washington University in St. Louis; the University of Iowa joined in the second year. The Mid-Atlantic branch began with the hub at Johns Hopkins and nodes at the University of Delaware, Drexel University, and the University of Maryland, Baltimore County; George Mason University and the University of Maryland, College Park joined soon after. The Math Alliance[6] is also a node for both branches. We serve as directors of INMAS, with support from associate directors Lukas Leisman of UIUC in the Midwest and Martin Lacasse of Johns Hopkins in the Mid-Atlantic. Leisman and Lacasse focus on connecting with companies and persuading them to create internship positions for Ph.D. students in our program.

Three distinct steps drive the annual INMAS workflow:

(i) Preparing student cohorts for internships
(ii) Curating internship positions
(iii) Placing students into internships.

Internship preparation involves both technical and professional components. For the technical side, INMAS students partake in four workshops on Python programming, statistics, machine learning, and group projects. These workshops initially occurred online due to the pandemic, but the program's current iteration is conducting most of them in person. In terms of professional skills, participants are coached on resume building, networking, and interview preparedness. INMAS associate directors serve as mentors to the students and gather insights into their strengths and employment preferences.

During the second, most innovative step in the INMAS workflow, we connect with prospective employers to "curate" internships. Companies with a scientific focus—such as remote sensing or carbon sequestration, for example—often already have some potential projects for applied mathematicians and are excited by the prospect of student interns. Other companies need some guidance; they might not usually consider hiring mathematical scientists but may be willing to try after hearing that the students can utilize data or undertake machine learning tasks.

We find the most success with small-to-medium-sized companies that do not have formal internship programs. We explain that INMAS will lower their in-house costs by using NSF funds to partially support the interns while simultaneously providing a pool of pre-screened candidates with relevant, applicable skills — therefore reducing the difficulty of finding qualified students. The internship component of INMAS thus enables companies without sizeable research staffs to nonetheless deploy modern mathematical tools.

Identifying and contacting potential host companies requires sustained effort throughout the year. We have done well with organizations that received grants through the U.S. government's Small

See **Internship Network** *on page 5*

In February 2023, the capstone "Project Workshop" of the Internship Network in the Mathematical Sciences (INMAS) united students from throughout the network for an intensive weekend of collaboration on problems in mathematical modeling and data science. Photo courtesy of the Department of Mathematics at the University of Illinois Urbana-Champaign.

[1] https://pi4.math.illinois.edu
[2] https://cse.umn.edu/ima/math-industry-boot-camp
[3] https://www.mitacs.ca/en
[4] https://new.nsf.gov/funding/opportunities/research-infrastructure-program
[5] https://inmas.us
[6] https://mathalliance.org

Computational Attorneys

Continued from page 1

bounds of existing legal systems in the relevant jurisdiction. At the same time, scientists must take appropriate measures to prevent the models from inventing seemingly plausible but nonexistent responses (i.e., hallucinations or confabulations). Retrieval-augmented generation [7]—as well as uncertainty quantification techniques like evidential deep learning—can help ensure stability and avoid confabulations.

Third, these models must be *provable*. In other words, their legal opinions or judgments need to stem from relevant laws and rules. Provability is a higher requirement than interpretability or explainability, as the models have to justify the correctness and fairness of each step of their reasoning process. Achieving this objective is crucial in the legal context and may involve techniques from abductive reasoning and neuro-symbolic inferencing.

Fourth, these models must be *communicable*. They should be capable of capturing subtle details and nuances in the instructions from fellow lawyers or legal clients. They should also be *teachable*, in that they keep learning from human demonstrations and feedback [2, 4]. The creation of an advanced natural language interface is essential for this specific purpose; such an interface must facilitate bidirectional communication, thereby allowing L^3Ms to learn from their interactions with humans and integrate new knowledge accordingly.

The development of functional computational attorneys is an exciting prospect that will likely play a central role in the future of the legal industry. It will enhance the efficiency of various legal services (such as conducting research, reviewing documents, preparing for depositions, drafting briefs, and creating contracts) and broaden access to legal assistance by dramatically reducing the associated costs. L^3Ms will comprise the foundation of this paradigm shift and serve as partners to human attorneys, providing expert support while the human attorneys retain overall control and responsibility (see Figure 1, on page 1).

Nevertheless, the use of computational attorneys is not without its potential challenges and unintended consequences. The aforementioned requirements are fairly demanding; if current L^3Ms do not meet them, additional safeguarding techniques will be necessary. We also need to ensure that L^3Ms function as tools that assist—but do not replace—human attorneys. Instead of jeopardizing employment in the legal workforce, the automation of simple and complex legal tasks should enable legal professionals to serve a much wider customer base. Moreover, information from computational attorneys cannot compromise a lawyer's legal and ethical responsibilities to provide accurate legal advice.

The legal jurisdictional systems in several regions have already begun to adapt to the transformations that result from generative AI technologies. In the U.S., judges have underscored an attorney's duty to ensure the accuracy of their legal statements and the robustness of their legal reasoning in their briefs [8]. Meanwhile, the European Bar Association has published guidelines[1] for best practices of lawyers in the era of ChatGPT. While ongoing research aims to establish the accountability of L^3Ms, the ultimate responsibility lies with human attorneys to meticulously review and rigorously verify the outputs of computational attorneys.

All in all, the future of legal AI promises to advance the legal profession in a multitude of dimensions. With further research and development, we can anticipate the rise of computational attorneys that not only execute mundane, low-level legal tasks with superhuman performance, but also take on complex, high-level legal challenges. This outcome would revolutionize the legal industry, provide more efficient legal services, and democratize access to justice.

[1] https://www.fbe.org/nt-commission-guidelines-on-generative-ai

References

[1] Bench-Capon, T., Araszkiewicz, M., Ashley, K., Atkinson, K., Bex, F., Borges, F., … Wyner, A.Z. (2012). A history of AI and law in 50 papers: 25 years of the international conference on AI and law. *Artif. Intell. Law*, *20*, 215-319.

[2] Casper, S., Davies, X., Shi, C., Gilbert, T.K., Scheurer, J., Rando, J., … Hadfield-Menell, D. (2023). Open problems and fundamental limitations of reinforcement learning from human feedback. Preprint, *arXiv:2307.15217*.

[3] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of law school. In *Findings of the association for computational linguistics: EMNLP 2020* (pp. 2898-2904). Association for Computational Linguistics.

[4] Christiano, P.F., Leike, J., Brown, T.B., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In *Advances in neural information processing systems 30 (NeurIPS 2017)*. Long Beach, CA: Curran Associates, Inc.

[5] Governatori, G., Bench-Capon, T., Verheij, B., Araszkiewicz, M., Francesconi, E., & Grabmair, M. (2022). Thirty years of artificial intelligence and law: The first decade. *Artif. Intell. Law*, *30*, 481-519.

[6] Katz, D.M., Bommarito, M.J., Gao, S., & Arredondo, P. (2023). GPT-4 passes the Bar exam. Preprint, *Social Science Research Network*.

[7] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in neural information processing systems 33 (NeurIPS 2020)* (pp. 9459-9474). Curran Associates, Inc.

[8] Merken, S. (2023, June 26). New York lawyers sanctioned for using fake ChatGPT cases in legal brief. *Reuters Legal*. Retrieved from https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22.

[9] Sartor, G., Araszkiewicz, M., Atkinson, K., Bex, F., van Engers, T., Francesconi, E., … Bench-Capon, T. (2022). Thirty years of artificial intelligence and law: The second decade. *Artif. Intell. Law*, *30*, 521-557.

[10] Surden, H. (2014). Machine learning and law. *Wash. Law Rev.*, *89*(1), 87.

[11] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., … Fedus, W. (2022). Emergent abilities of large language models. *Trans. Mach. Learn. Res.*

[12] Yu, F., Quartey, L., & Schilder, F. (2023). Exploring the effectiveness of prompt engineering for legal reasoning tasks. In *Findings of the association for computational linguistics: ACL 2023* (pp. 13582-13596). Toronto, Canada: Association for Computational Linguistics.

Frank Schilder is a senior research director at Thomson Reuters Labs, where he leads a team of researchers and engineers who explore new machine learning and artificial intelligence (AI) techniques to create smart products for legal natural language processing problems. Prior to joining Thomson Reuters, Schilder was an assistant professor in the Department of Informatics at the University of Hamburg. Dell Zhang currently leads the Applied Research team at Thomson Reuters Labs in London. He was formerly a tech lead manager at ByteDance AI Lab and TikTok UK; a staff research scientist at Blue Prism AI Labs; and a reader in computer science at Birkbeck, University of London.

SIAM Conference Formats Continue to Innovate

By Richard Moore

At the end of each calendar year, I take the opportunity to reflect on SIAM conferences — how they've fared in the recent past and where they're headed in the near future. Luckily for me, SIAM Vice President for Programs Jim Nagy and SIAM President Sven Leyffer have already summarized and reflected upon previous conference configurations in their September *SIAM News* article.[1] I will therefore focus on the year ahead.

As Jim and Sven explained, post-conference survey responses from SIAM members have made two facts abundantly clear. First, virtual and hybrid formats are important avenues that increase the accessibility of SIAM conferences for those who cannot travel to a meeting site due to a variety of circumstances, ranging from illness to visa restrictions and budgetary considerations (both financial and carbon).[2] Second, these formats do not yet provide a comparable experience to the energy and sense of community of an in-person SIAM conference. Moreover, hybrid components can sometimes detract from the in-person setting when, for example, the audience at a remote plenary lecture feels disengaged from the speaker. As the COVID-19 pandemic continues to wane, we have observed diminishing remote participation in hybrid conferences: from 41 percent at the inaugural SIAM Conference on Mathematics of Data Science[3] in October 2022 to just four percent at the ACM-SIAM Symposium on Discrete Algorithms[4] in January 2023 and seven percent at the SIAM Conference on Applied Algebraic Geometry[5] in July 2023. Meanwhile, the SIAM Conference on Computational Science and Engineering[6] in February-March 2023, which took place strictly in person in Amsterdam, the Netherlands, attracted a total of 2,054 attendees — the highest-ever number of registrants at an in-person SIAM meeting.

So, what is the right balance between the accessibility of virtual and hybrid formats versus the unique experience of physically attending a SIAM conference? We don't claim to have the answer but remain committed to building on all that we have learned over the past few years.

SIAM's next experiment with virtual formatting will occur at the 2024 SIAM Annual Meeting[7] (AN24), which will take place in July. AN24 will consist of two parts: an in-person event from July 8-12 at the Spokane Convention Center in Spokane, Wash., and an online component from July 18-20. Each of the three online days will begin and end at different times to better accommodate SIAM's three largest membership bases in North America, Europe, and Asia. The online event will include the following elements:

- Timed streaming of recorded plenary and prize talks from the previous week's in-person meeting
- A live question-and-answer session with the speaker and moderator (depending on their availability) after each streamed plenary or prize talk; the talks and recorded discussions will then be available on demand
- Strictly online-only minisymposia and contributed sessions that will be recorded and made available on demand after the sessions conclude; speakers may be asked to send recordings of their talks in advance, but they will be present in the online forum to chat about their work with other attendees
- Roundtable discussions that online participants can organize themselves
- Casual networking events.

Everyone who submits their work to AN24 will choose whether they wish to present online or in Spokane. Contributed speakers and organizers of online minisymposia must provide their time zones so that talks can be scheduled at a reasonable local hour. Once a speaker is slated to present virtually or in person, they will not be permitted to change their assigned format. All accepted speakers need to register by the early registration deadline or risk removal from the conference program.

The 2024 SIAM Conference on Applied Mathematics Education[8] (ED24) and the 2024 SIAM Conference on Discrete Mathematics[9] (DM24) will be co-located with AN24. Both of these meetings will operate strictly in person, as they have done historically. We regard the exploration of new formats as an experiment and intend to learn more about the relevant challenges and concerns before recommending changes to the SIAM Activity Groups (SIAGs) regarding their affiliated conferences. However, a speaker from any SIAG community who wishes to present remotely can submit to the online component of AN24, as the SIAM Annual Meeting welcomes contributions from all SIAGs.

Rather than implement a complicated fee structure like we did for hybrid conferences in 2022, SIAM now favors a single registration fee for each category (student/nonstudent, member/nonmember, etc.). A single registration will provide access to in-person attendance at AN24, ED24, and DM24, as well as online attendance at AN24; the sole exception to this rule is a one-day registration that will only grant in-person access for the selected day. This model allows participants as much flexibility as possible and ensures that online attendees can still benefit from the conference without the burden of travel expenses. If a speaker is invited to present at an online-only minisymposium due to constraints of the session organizer, that individual can still choose to attend the in-person component in Spokane. Importantly, SIAM will continue to offer the outreach member[10] rate that we implemented for hybrid conferences in 2022; outreach member registrants are now able to attend and present in person for a substantially reduced rate.

To conclude, I would like to thank the National Science Foundation and U.S. Department of Energy for their generous support of SIAM conferences, and briefly mention some exciting new conference features for 2024. Onsite badge printing will make its debut at AN24 to help mitigate long registration lines. Also forthcoming is the implementation of pre-conference minicourses, the first of which is expected to take place in Atlanta, Ga., on October 20 — the day before the start of the 2024 SIAM Conference on Mathematics of Data Science.[11] In addition, attendees will once again be able to take advantage of the convenient "nonmember + join" option that bundles nonmember conference registration with a regular SIAM membership at all 2024 meetings.[12] And we will continue to provide travel support for SIAM conference attendees via our Student Travel Awards[13] and Early Career Travel Awards,[14] which rely on funds from grants, donors, and SIAM's operational budget. Finally, don't forget to stop by the membership desk at the next SIAM conference to grab a new SIAM sticker for your laptop and chat with our friendly membership staff!

As always, I am truly indebted to the conference team at SIAM—led by Conference Director Lisa Dyson—for their professionalism and commitment to excellence in delivering world-class conferences month after month. Also, the *SIAM News* article by Jim Nagy and Sven Leyffer that I previously referenced fails to mention the countless meetings, consultations, and decisions that confronted Jim when COVID-19 struck two months into his first term as Vice President for Programs. His farsighted vision—as well as support from the SIAM Board of Trustees, SIAM President Sven Leyffer, and Past Presidents Susanne Brenner and Lisa Fauci—have allowed SIAM conferences to weather the storm of the pandemic and emerge with renewed vigor and resilience. We look forward to seeing you, either in person or online, at an upcoming SIAM conference!

Richard Moore is the Director of Programs and Services at SIAM.

Richard Moore, SIAM's Director of Programs and Services, enjoys the Upside Down Amsterdam—the largest social media museum in Amsterdam, the Netherlands—after the 2023 SIAM Conference on Computational Science and Engineering earlier this year. Here he poses in the Mondriaan Room, which honors the work of Dutch painter Piet Mondriaan. Photo courtesy of the Upside Down Amsterdam.

[1] https://sinews.siam.org/Details-Page/reflecting-on-and-reinventing-siam-conferences
[2] https://sinews.siam.org/Details-Page/forward-looking-panel-at-ds23-envisions-the-future-of-conferences
[3] https://www.siam.org/conferences/cm/conference/mds22
[4] https://www.siam.org/conferences/cm/conference/soda23
[5] https://www.siam.org/conferences/cm/conference/ag23
[6] https://www.siam.org/conferences/cm/conference/cse23
[7] https://www.siam.org/conferences/cm/conference/an24
[8] https://www.siam.org/conferences/cm/conference/ed24
[9] https://www.siam.org/conferences/cm/conference/dm24
[10] https://www.siam.org/membership/join-siam/individual-members/outreach-membership
[11] https://www.siam.org/conferences/cm/conference/mds24
[12] https://www.siam.org/conferences/calendar
[13] https://www.siam.org/conferences/conference-support/siam-student-travel-awards
[14] https://www.siam.org/conferences/conference-support/siam-early-career-travel-awards

Internship Network

Continued from page 3

Business Innovation Research and Small Business Technology Transfer programs.[7] Ph.D. alumni from INMAS node universities who currently work in industry serve as another good source of connections, as do companies in campus research parks and regional business incubators.

During the internship placement step, we attempt to solve the *stable matching problem* in a collaborative way. Students read project descriptions from the companies and communicate their preferred choices to our associate directors. The associate directors then send a subset of qualified students to each company for consideration. The companies review the resumes, conduct interviews, and typically make an offer to one or more students.

Key Takeaways of INMAS

INMAS' network structure helps to build a community of mathematical scientists with similar interests who will support each other in their career journeys. It also allows us to pool resources so the directors can train students at scale. Perhaps most importantly, INMAS fills a gap that exists within every department in the network: a lack of personnel who work specifically to connect students with internships.

The direct connection between students and employers is critical to INMAS' success. Simply training students is not enough because most mathematics graduate students need guidance when securing their first internships. In fact, INMAS participants who apply independently for positions—as we encourage them to do—often submit hundreds of applications to no avail. But host companies have been uniformly impressed with the contributions of their INMAS interns, who serve as firsthand proof that industry can benefit from looking beyond the traditional recruiting pool of engineering students. Students who complete an internship through INMAS are much better positioned to independently obtain another internship or full-time employment in subsequent years.

The culture change in Ph.D. programs at node universities is also important, as students who enjoy their internships talk with peers and normalize the (perfectly normal) goal of a career in industry. This type of attitude shift is long overdue in many mathematics departments.

Over the past three years, INMAS has provided training to 200 students. 33 of these students found internships independently, while INMAS placed 107 students into internship positions (the remaining students chose other paths or deferred internships to the following year). In 2023 alone, 32 different companies throughout the U.S. hosted INMAS interns.

Programs with Similar Missions

Mitacs, which served as our inspiration, is a significant contributor to workforce development in Canada. Established in 1999, its major internship program—Mitacs Accelerate[8]—now places more than 15,000 students in positions each year [1]. Like with INMAS, the partner companies receive a subsidy for each project. Mitacs' reach is broad and includes nearly all disciplines; their funding comprises more than CAD $200 million per year and comes from federal and provincial government sources.

The Erdös Institute,[9] which originated at Ohio State University, also provides training and job connections for Ph.D. students in the sciences. The Institute collaborates with both universities and companies; students from partner institutions receive training in several key areas of coding and data science that are valuable in the job market, and partner companies have access to a talent pool for full-time hires after graduation. Unlike Erdös, INMAS focuses on internship placement *prior* to graduation.

Over the last decade, bootcamp-type programs that prepare new and recent Ph.D.s for data science and engineering roles have grown in popularity. Organizations such as the Insight Fellows Program,[10] The Data Incubator,[11] and NYC Data Science Academy[12] all offer this type of training. The business model for such programs is continually evolving, and participants must generally pay a substantial tuition fee.

In recent years, NSF also partnered with the Oak Ridge Institute for Science and Education to create the Mathematical Sciences Graduate Internship Program.[13] Participating students spend the summer in Department of Energy labs and work under the supervision of laboratory scientists on projects that often have an academic flavor. In contrast, INMAS interns typically tackle projects that impact the immediate future of the host company.

Students tackle open-ended projects in mathematical modeling and data science during the capstone "Project Workshop" of the Internship Network in the Mathematical Sciences (INMAS), which took place in February 2023. Photo courtesy of the Department of Mathematics at the University of Illinois Urbana-Champaign.

Preparing All Students for Internships

INMAS centralizes training and placement to effectively handle a large volume of students. This approach—while certainly efficient—requires a tremendous amount of effort to coordinate the many interdependent parts and their logistical complexities. However, we believe that individual university departments can achieve similarly good results on a smaller scale.

For instance, students can obtain useful technical training through several avenues:

- Taking a Python programming course on campus or online (departments can reimburse students for any costs)
- Taking an intermediate statistics or statistical learning course, or an online course with certification
- Enrolling in a machine learning course (through another department if necessary)
- Forming a team and entering a Kaggle competition.[14]

In addition, students can hone their professional skills via certain steps:

- Working with campus career services and faculty members and utilizing online resources to prepare a resume and LinkedIn profile
- Reading SIAM's *BIG Jobs Guide*,[15] a practical career book for students in the mathematical sciences [2]
- Studying training material about elevator pitches and behavioral interviews from sites like Glassdoor and Indeed
- Practicing interview skills with peers.

Departments can create internship opportunities for their students by appointing a faculty member to develop relationships with relevant companies in the following ways:

- Inviting alumni who work in industry to speak on campus about their career journeys and roles at their current companies
- Connecting with university-run industrial parks or tech incubators
- Cold-calling or emailing local employers, particularly those with science Ph.D.s on staff
- Encouraging students to also apply for internships directly.

Departments should recognize the efforts of this designated faculty member. Their work with graduate students and companies is not a voluntary "extra" — it is an important contribution to both the department and the future of the profession.

Finally, SIAM provides a wealth of information for job seekers on its Career Resources page[16] and Internships page,[17] through the SIAM Job Board,[18] and via its regular career fairs.[19] The SIAM Industry Committee[20] also hosts panel discussions during which SIAM members who work in industry share their experiences and offer advice to attendees.

Of course, there is always more to be done. We encourage readers to contact us with questions and ideas at **laugesen@illinois.edu** or **fsantos9@jhu.edu**, and please visit the INMAS website[21] for more information. We look forward to hearing from you!

View the online version of this article for student photos and testimonies about their individual experiences.

References

[1] Adem, A. (2019). Mitacs: Mathematical roots for innovation in Canada. *Notices Am. Math. Soc.*, *66*(8), 1290-1293.

[2] Levy, R., Laugesen, R., & Santosa, F. (2018). *BIG jobs guide: Business, industry, and government careers for mathematical scientists, statisticians, and operations researchers*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Richard Laugesen has taught at the University of Illinois Urbana-Champaign since 1997, where his research focuses on partial differential equations and spectral theory. As a former Director of Graduate Studies, he promotes mathematical careers in industry and government and works toward a future in which more women and minority students pursue graduate-level mathematics. Fadil Santosa is the Yu Wu and Chaomei Chen Department Head of Applied Mathematics and Statistics at Johns Hopkins University. He was formerly a professor of mathematics at the University of Minnesota and served as director of the Institute for Mathematics and its Applications from 2008 to 2017. Santosa works in several areas of applied mathematics, including inverse problems, optimal design, and optics.

[7] https://www.sbir.gov/about
[8] https://www.mitacs.ca/en/programs/accelerate
[9] https://www.erdosinstitute.org
[10] https://insightfellows.com
[11] https://www.thedataincubator.com
[12] https://nycdatascience.com
[13] https://orise.orau.gov/nsf-msgi
[14] https://www.kaggle.com/competitions
[15] https://my.siam.org/Store/Product/viewproduct/?ProductId=29783110
[16] https://www.siam.org/careers/resources
[17] https://www.siam.org/careers/internships
[18] https://jobs.siam.org
[19] https://www.siam.org/careers/resources/siam-career-fairs
[20] https://www.siam.org/about-siam/committees/industry-committee
[21] https://inmas.us

Trucking, Surveillance, and Applied Mathematics

BOOK REVIEW
By Ernest Davis

Data Driven: Truckers, Technology, and the New Workplace Surveillance. *By Karen Levy. Princeton University Press, Princeton, NJ, December 2022. 240 pages, $33.00.*

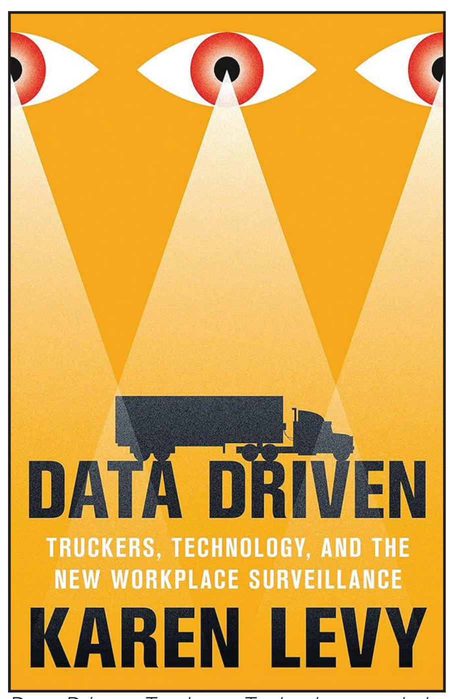
Long-haul trucking is a demanding, dangerous, and poorly compensated job. When adjusted for inflation, the average yearly income for truck drivers declined from roughly $110,000 in 1980 to $47,000 in 2020. In 2019, trucking had the sixth highest fatality rate among U.S. occupations and the highest rate of serious (but not fatal) injuries. The turnover of truck drivers at large firms is close to 100 percent each year.

Truck driving is explicitly *not* covered by the U.S. Fair Labor Standards Act, which requires overtime pay and other forms of worker protection against exploitation. And unlike most blue-collar jobs, truckers are generally not paid by the hour; instead, they are paid by the miles that they drive. This type of payment structure creates perverse incentives, in that workers are tempted to prioritize speed over safety and spend more hours driving and fewer hours sleeping than is otherwise advisable. Yet despite its image as a haven for free-spirited scofflaws on the open road, trucking is in fact very tightly regulated.

For example, truckers are prohibited from driving more than 10 hours per day. Starting in the 1930s, drivers were required to keep written logbooks of their activities that were subject to regular inspection. These logbooks invited many forms of falsification, and cheating was rampant. In the late 1980s, large trucking companies began to utilize electronic logging devices (ELDs): small instruments that are hardwired to a truck's engine and record a log of its engine status, location, and mileage. In 2012, Congress passed a law that mandated the use of ELDs; the law took effect in 2017. In her recent book, *Data Driven: Truckers, Technology, and the New Workplace Surveillance*, Cornell University sociologist Karen Levy examines the effect and implications of these devices on members of the trucking industry, overall job surveillance, and society at large.

The Federal Motor Carrier Safety Administration's original requirements for ELDs were quite vague and provided only minimal specification of the data that was eligible for collection. As such, the many different models of available ELDs addressed the distinct objectives of various markets. On one end of the spectrum were ELDs that continually reported as much data as possible—i.e., a truck's status, the process of driving, and even biometric information—to a corporate control center that could immediately communicate its demands to the driver. Large trucking companies generally preferred these models. At the other end of the spectrum were ELDs that appealed to the drivers themselves and only incorporated the minimum number of legally required features; though not explicitly stated, these ELDs often allowed for manipulation by the drivers.

Who benefited from ELDs? For one, ELD manufacturers now had a captive market. Large trucking companies also typically favored ELDs because they offered tighter control of their operations; in fact, these companies had largely deployed ELDs in their trucks long before the government mandate. Many smaller trucking companies, however—which constitute a large fraction of all trucking activity—generally disliked the requirement; ELDs were expensive, complicated the interactions between companies and truckers, and did not supply particularly useful information (indeed, many small companies suspected that larger corporations lobbied for the ELD mandate in order to outcompete them).



Data Driven: Truckers, Technology, and the New Workplace Surveillance. By Karen Levy. Courtesy of Princeton University Press.

Truckers generally hated ELDs and viewed them as intrusive. Unsurprisingly, drivers soon found all kinds of ways to circumvent the devices — from crudely smashing them with hammers to disabling them or avoiding their surveillance. The ELD-based instructions from control centers were usually annoying and often oblivious to the actual situation at hand. Nevertheless, Levy remains reasonably cautious in making specific claims as to the extent of overall harm that ELDs ultimately caused truckers.

One might suppose that ELDs would appeal to government inspectors, who no longer had to worry about logbook falsification. But in actuality, their response was not clear-cut. Because the inspectors were initially unfamiliar with the many different ELD models, they had to rely on truckers to explain the devices and often simply took the truckers' word on the ELD reports. In many cases, inspectors and truckers found themselves in a sort of alliance against the electronic devices.

The benefits of the ELD directive for the public are also unclear. The mandate sought to make trucking safer by preventing truckers from driving when fatigued, but there is reason to think that trucking accidents actually *increased* after it was imposed. Any safety gains from shorter driving stints were immediately lost when truckers drove faster and more recklessly to finish their routes within the allotted 10 hours.

Data Driven discusses at some length the fact that the smooth operation of modern society depends significantly on rule breaking. Levy is not at all doctrinaire on issues of regulation or technology. She does not argue that the world would be a better place if workplace regulations were abolished and the unfettered free market could operate without hindrance, or that the effects of regulation are inevitably dominated by the "law of unintended consequences;" nor does she contend that all technology or even all surveillance mechanisms merely serve to increase the power of corporations over workers. Levy maintains that the formulation and enforcement of regulations *and* the design and deployment of technology should be done *wisely*, with careful monitoring of actual consequences and a realistic and sympathetic view of the complex, multifarious realities of the situation and people involved. She also advocates for specific reforms in the trucking industry to improve the working conditions of truckers at some cost to trucking corporation profits, such as paying truckers by the hour rather than the mile.

One chapter of *Data Driven* addresses the possible future impact of artificial intelligence (AI) technology, particularly self-driving vehicles. Levy's assessment is cautiously optimistic, much like that of Stanford University economist Erik Brynjolfsson. Given the political will, Levy believes that scientists can gradually and effectively introduce AI technology to society in a manner that avoids sudden disruptions and ultimately improves the efficiency, safety, and working conditions of the trucking industry.

Levy primarily carried out her research by interviewing the various actors at play: truckers, inspectors, corporate administrators, and government regulators. Her book features an interesting discussion about the techniques that she used to encourage her interlocutors to speak candidly, as well as an introspective reflection of the way in which her own personal characteristics as a young, educated, white woman helped and hindered the conversations. Levy was also visibly pregnant for a time, which had both pros and cons; people—particularly women—become more sympathetic and forthcoming, but she could no longer conduct interviews over a beer.

So, what is the role of applied mathematics in this field of inquiry? *Data Driven* contains essentially no math. There are numbers, certainly, but nothing that requires more than a basic understanding of simple arithmetic. Levy does not conduct any statistical analyses or computations of levels of confidence, confidence intervals, or Bayesian posteriors.

Of course, ELD technology builds on mathematics, mathematical physics, and the theory of computation in well-known ways. The internal computational mechanism of ELDs draws upon digital computation that dates back to the work of Claude Shannon and his predecessors, and the communication mechanisms stem from the theory of analogue waves by Joseph Fourier and those that came before him. A narrow but significant pathway to differential geometry and tensor theory even exists, given that ELDs use the Global Positioning System (GPS) for location and clocks in GPS satellites must account for the gravitational effects that are posited by general relativity. Behind the scenes, statisticians in the policy and planning departments of both corporations and government agencies are undoubtedly deploying statistical packages and decision support software that squeeze all possible information out of the vast quantities of data that ELDs constantly generate in millions of U.S. trucks; they presumably use this information to inform corporate strategies and government policies.

Regardless, it seems to me that the more important question that Levy's work poses for *SIAM News* readers is whether and how mathematics can help achieve society's best goals. "As digital technologies flourish across social life, they stand poised to reorient how people relate to institutions, to each other, and to themselves," Levy writes. "Often, the best way to think about technological change is not to focus solely on the technology, but to strengthen the social institutions and relations that surround it. Only by doing so can we ensure that digital technologies become part of a vibrant social order that protects workers and promotes human dignity." I hope and believe that mathematical thinking could be a part of this process.

Ernest Davis is a professor of computer science at New York University's Courant Institute of Mathematical Sciences.

How We Can All Play the High-performance Game: From Laptops to Supercomputers

By Michael Bader, Hatem Ltaief, Lois Curfman McInnes, and Rio Yokota

The changing landscape of heterogeneous computing offers both unprecedented power and unprecedented complexity [7]. Given the associated challenges and opportunities, we discuss the necessity of international collaboration to develop open-source portable scientific software ecosystems, including the required functionality for advanced scientific computing. Such collaboration will enable all members of our community to fully exploit the computational power of hybrid machines—from laptops and desktops to clusters, supercomputers, and beyond—and advance scientific discovery.

In a recent letter to the editor in *SIAM News*,[1] Nick Trefethen called attention to the divergence in computing performance on high-end supercomputers versus everyday laptop and desktop machines [12]. His article raises important considerations about the evolving landscape of computer architectures and implications for the future of computational science and high-performance computing (HPC). As current officers of the SIAM Activity Group on Supercomputing[2] (SIAG/SC), we respond by considering heterogeneous computing architectures that are now readily available at all scales of computing, including on cell phones, laptops, desktops, and clusters. Heterogeneous designs—typically combinations of both central processing units (CPUs) and graphics processing units (GPUs) [11]—offer the potential of unprecedented computing power in a given computing environment (say, on a desktop); however, their extraordinary complexity means that most people cannot fully take advantage of that performance in their own local computing environments.

Trefethen points out that the trend in computational power of the average desktop machine (shown by the solid pink line in Figure 1) has advanced several orders of magnitude less than that of the largest machines on the TOP500 list. Indeed, most typical users of laptops and desktops readily write CPU-only code and achieve the level of performance in the figure.

Architectural Trends: Processors are Getting Faster, But Only With Parallel Codes

In the quest towards petascale and exascale computing, energy efficiency requirements have forced architectures to extreme parallelism and complicated memory hierarchies. Even mainstream laptops and desktops are now parallel hybrid architectures, including both CPUs and GPUs. Figure 1 demonstrates that when we fully leverage modern CPUs and GPUs, the growth of laptop and desktop computing power actually aligns with the growth of extreme-scale computing. However, the traditional approach to computing—which only employs the simple-to-use CPUs and omits the more difficult GPUs—leaves behind an enormous fraction of untapped potential in the machines. A huge challenge therefore remains: Given the architectural complexity of these hybrid systems and the limited performance-portable, open-source software that is tailored for these new architectures, how can we all play the HPC game?

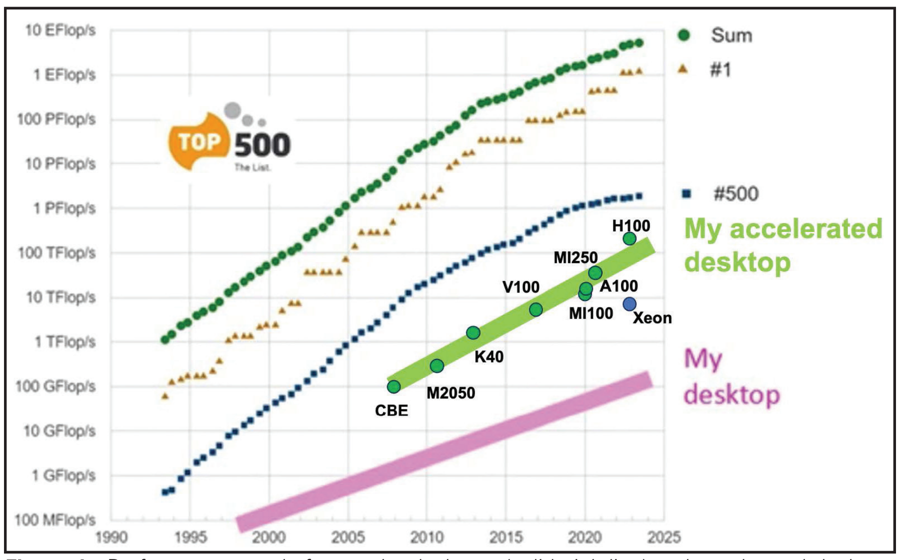


Figure 1. *Performance trends for regular desktops (solid pink line) and accelerated desktops (solid green line). A fixed budget no longer supports this trend. An NVIDIA H100 graphics processing unit costs roughly $30,000; when we talk about performance, we are probably assuming that the system is actually a high-end server. Figure augmented from a graph that appeared in [12], which was augmented from a TOP500 Project graph.*

Extracting the Full Potential of CPUs is Getting Harder

The high-end Intel Xeon Platinum 9282 has a theoretical peak of 9.3 TFlop/s (see Figure 1). This performance can be achieved only if the code can utilize the 56 cores and AVX-512 fused multiply-add instructions with a single instruction, multiple data vector width of eight doubles. If the code is not vectorized, users get 1/8 of the performance; if it is not multithreaded, they get another 1/56 of that. In practice, many issues—such as the algorithm's arithmetic intensity and the spatial/temporal locality of the implementation—further reduce performance. Sole reliance on compilers to automatically solve all of these issues has been unsuccessful. Furthermore, porting code to GPUs is not the only challenge; both CPUs and GPUs may suffer from inadequate numerical algorithms and software ecosystems.

SOFTWARE AND PROGRAMMING

Deep Learning Inference Drives Hardware Architecture Design

Deep learning applications can tolerate extremely low levels of precision, and GPU vendors are taking advantage of this fact. While NVIDIA H100 GPUs achieve 134 TFlop/s in FP64 with Tensor Cores, they have a theoretical peak of about 2 PFlop/s for code that can fully utilize the FP16/BF16 sparse Tensor Cores, and close to 4 PFlop/s for code that can use FP8/INT8. Another bifurcation will evolve here if the scientific computing community does not exploit this extra computing power.

Techniques exist to recover double-precision accuracy via low-precision Tensor Cores at the matrix multiplication level [9, 10] and the solver level [1, 4]. But at the moment, such techniques are primarily exploited for certain dense/sparse linear algebra operations. In order for a wider range of scientific applications to benefit from these architectures, we need cross-cutting research in three key areas: (i) Theory and guidelines to profit from mixed precision while maintaining the fidelity of results [6], (ii) simulation software to improve performance portability, and (iii) robust and trustworthy software ecosystems to encapsulate functionality [8].

The Deep Learning Community Benefits From Algorithmic Uniformity

Scientific computing no longer drives supercomputing architectures. GPUs were originally developed to solve a few well-defined tasks in the realm of games and interactive graphics computing. Now they serve the deep learning community by championing the forward and backward propagation of neural networks. This circumstance allows hardware manufacturers to aggressively optimize their architectures for a very narrow set of algorithms. Because of energy and cost constraints, processors can no longer simultaneously be general purpose *and* high performance. Neural network architectures have coevolved with GPUs; transformers (types of deep learning models) can now extract a large portion of the Tensor Core's theoretical peak Flop/s.

Architectural Features Drive Progress in High-performance Scientific Computing

The diversity of algorithms that are needed to address the broad range of scientific computing applications complicates the unified coevolution approach (see Figure 2, on page 8). Does this mean that the scientific computing community will not be able to fully embrace the hardware evolution? We must redesign current algorithms, devise new algorithms, and refactor and redesign code to fully leverage the fast, low-precision hardware units and other features of heterogeneous architectures for scientific computing. A variety of groups within the international community have been pushing towards exascale computing with advances in algorithms and software [2, 3]. Because research gains that are driven by these extreme-scale systems tackle the on-node functionality that manifests throughout all scales of computing, this work also provides a foundation for performance improvement across desktops and clusters.

An Urgent Need for New Research and Scientific Software Ecosystems

The scientific computing community faces an urgent need for research on numerical algorithms that exploits the features of heterogeneous architectures, including mixed precision (wherever applicable), complex memory hierarchies, and massive parallelism. Equally important is the development of open-source community software ecosystems that encapsulate this functionality for ready use by everyone. A combined approach of algorithmic advances and international collaboration on robust and trustworthy scientific software ecosystems will enable us to counter bifurcation and fully exploit the computing power of new heterogeneous architectures at all scales of computing.

Get Involved

Addressing the challenges of next-generation algorithms and software requires a wide range of expertise from the international scientific computing community, including members of multiple SIAGs that address relevant topics: SIAG/SC,[3] SIAG on Computational Science and Engineering,[4] SIAG on Data Science,[5] SIAG on Linear Algebra,[6] SIAG on Applied Mathematics Education,[7] and SIAG on Equity, Diversity, and Inclusion.[8] Also valuable are topical groups that focus on imaging science, life sciences, and geosciences, as well as partnerships with the IEEE Computer Society's Technical Community on Parallel Processing[9] and the Association for Computing Machinery's Special Interest Group on High Performance Computing.[10]

Interested in high-performance scientific computing? We encourage you to join SIAG/SC to engage with a vibrant community of scientists who consider a broad range of topics in numerical algorithms and computer architectures that directly contribute to high-performance computer systems. SIAG/SC promotes the exchange of ideas by focusing on the interplay of analytical methods, numerical analysis, and efficient computation.

Interested readers should consider attending the 2024 SIAM Conference on Parallel Processing for Scientific Computing[11] (PP24), which will take place from March 5-8, 2024, in Baltimore, Md. Featuring an

See **High-performance Game** *on page 8*

[1] https://sinews.siam.org/Details-Page/a-bifurcation-in-moores-law
[2] https://www.siam.org/membership/activity-groups/detail/supercomputing
[3] https://siag-sc.org
[4] https://www.siam.org/membership/activity-groups/detail/computational-science-and-engineering
[5] https://www.siam.org/membership/activity-groups/detail/data-science
[6] https://www.siam.org/membership/activity-groups/detail/linear-algebra
[7] https://www.siam.org/membership/activity-groups/detail/applied-mathematics-education
[8] https://www.siam.org/membership/activity-groups/detail/equity-diversity-and-inclusion
[9] https://tc.computer.org/tcpp
[10] https://www.sighpc.org
[11] https://www.siam.org/conferences/cm/conference/pp24

Another Year of Gratitude at SIAM

By Abby Addy

It's hard to believe that the end of my second year at SIAM is already approaching. I've been so fortunate to become better acquainted with our community over the past year and am excited to share some of the accomplishments of SIAM's Development and Corporate Relations team in 2023. Our continued growth is facilitated by the generosity of members and friends like you!

Development and Corporate Relations recently assumed responsibility of SIAM's conference sponsorships and exhibitors (which were previously housed in the Marketing and Communications Department). This change allows us to take an increasingly holistic approach when building relationships with sponsors and exhibitors, which will in turn provide more robust support for SIAM conferences and their attendees. If you work for an organization that you would like to see represented at a future SIAM conference, please reach out to **sponsorship@siam.org** for more details.

We've also established several impactful new gift funds this year. Two such examples are the *Life Sciences and Dynamical Systems Travel Fund*[1]—which was made possible by the generosity of Simone Bianco of the Altos Labs Bay Area Institute of Science—and the *Math Modeling Teacher Support Fund*[2] in honor of Michelle Montgomery, who recently retired from her role as the MathWorks Math Modeling Challenge[3] Program Director after a 35-year career at SIAM. 2023 has also seen continued contributions to funds like the *Postdoctoral Support Fund*,[4] which awarded its first grants this year through the SIAM Postdoctoral Support Program.[5] I wish to extend my sincere thanks to everyone who has donated to SIAM's gift funds and helped foster our efforts. I'd like to particularly highlight donors to the *SIAM General Fund*, which provides unrestricted support to the Society's highest-priority projects.

In partnership with the leadership of the SIAM Activity Group on Optimization[6] (SIAG/OPT), we have been undertaking a fundraising campaign to establish an endowment for the *SIAG/OPT Test of Time Award*.[7] This award recognizes an outstanding piece of work that has had a significant and sustained influence on the field of optimization. If you or someone you know would like to support the award's endowment, please reach out! Raising the endowment has been a genuine group effort, and I'm truly impressed by the way in which SIAM members and friends have come together.

In late 2022, SIAM announced the establishment of the Renata Babuška Prize—which recognizes high-quality interdisciplinary work that targets any aspect of modeling and numerical solution of a specific engineering or scientific application—in response to a generous donation by longtime SIAM member Ivo Babuška. When Ivo passed away in April,[8] the name of the award was formally changed to the *Ivo & Renata Babuška Prize*[9] in his honor. SIAM is grateful to Ivo for his many contributions to the field and proud to carry on his impressive legacy through this biennial prize.

Here I've provided just a small sample of the projects that SIAM's Development and Corporate Relations team has pursued this year, but they all inspire the same sentiment: thankfulness for everything that you do to support SIAM. I hope that I've had a chance to thank you personally; if not, I hope that 2024 brings more opportunities to connect! I remain encouraged by and grateful for the generosity of spirit that sets SIAM apart — from the welcoming atmosphere at conferences and events to the donations that sustain our important work. I feel privileged to be a part of SIAM.

Additionally, please save the date for **July 11, 2024**, when SIAM will host a donor recognition event during the 2024 SIAM Annual Meeting[10] in Spokane, Wash. This gathering will celebrate the generosity of SIAM members and the initiatives that your support has made possible. I'm looking forward to this special chance to bring our donors together and can't wait to see you there.

If you're considering a philanthropic donation to SIAM this year but haven't yet made your gift, there's still time! Gifts of all sizes have tremendous impact, and you can direct your contribution to your preferred area of support. To make a gift, visit SIAM's online giving page[11] or send a check payable to SIAM to the following address:

SIAM Development
c/o Abby Addy
3600 Market Street, 6th Floor
Philadelphia, PA 19104

Please don't hesitate to contact me at **(267)-648-3529** or **aaddy@siam.org** with any questions or comments, for assistance with your contribution, or just to chat. I look forward to connecting with you and wish you and your families a happy and healthy holiday season!

Abby Addy is the Director of Development and Corporate Relations at SIAM.

[1] https://sinews.siam.org/Details-Page/siam-establishes-the-life-sciences-and-dynamical-systems-travel-fund
[2] https://sinews.siam.org/Details-Page/siam-establishes-the-math-modeling-teacher-support-fund
[3] https://m3challenge.siam.org
[4] https://sinews.siam.org/Details-Page/siam-establishes-postdoctoral-support-fund
[5] https://www.siam.org/students-education/programs-initiatives/siam-postdoctoral-support-program
[6] https://www.siam.org/membership/activity-groups/detail/optimization
[7] https://www.siam.org/prizes-recognition/activity-group-prizes/detail/siag-opt-test-of-time
[8] https://sinews.siam.org/Details-Page/remembering-ivo-babuška
[9] https://www.siam.org/prizes-recognition/major-prizes-lectures/detail/ivo-and-renata-babuška-prize
[10] https://www.siam.org/conferences/cm/conference/an24
[11] https://www.siam.org/donate

Abby Addy, SIAM's Director of Development and Corporate Relations (left), converses with SIAM Membership Manager Paula White during the "Mobilizing the Power of Diversity: Celebrating 25 Years of EDGE" event, which took place in October 2023 at Bryn Mawr College. SIAM was a sponsor of this two-day gathering. Photo courtesy of Paola Nogueras and the EDGE Program.

High-performance Game

Continued from page 7

exciting program that highlights the newest advances in the field, PP24 is a great opportunity for attendees to learn about and discuss the latest happenings in high-performance scientific computing. Please join us!

A slightly amended version of this article[12] published on the SIAG/SC website on November 8th.

[12] https://siag-sc.org/hpc-for-all.html

References

[1] Abdelfattah, A., Anzt, H., Boman, E.G., Carson, E., Cojean, T., Dongarra, J., … Yang, U.M. (2021). A survey of numerical linear algebra methods utilizing mixed-precision arithmetic. *Int. J. High Perform. Comput. Appl.*, *35*(4), 344-369.

[2] Alam, S.R., McInnes, L.C., & Nakajima, K. (2022). IEEE special issue on innovative R&D toward the exascale era. *IEEE Trans. Parallel Distrib. Syst.*, *33*(4), 736-738.

[3] Anzt, H., Boman, E., Falgout, R., Ghysels, P., Heroux, M., Li, X., … Yang, U.M. (2020). Preparing sparse solvers for exascale computing. *Phil. Trans. R. Soc. A: Math. Phys. Eng. Sci.*, *378*(2166), 20190053.

[4] Haidar, A., Bayraktar, H., Tomov, S., Dongarra, J., & Higham, N.J. (2020). Mixed-precision iterative refinement using tensor cores on GPUs to accelerate solution of linear systems. *Proc. R. Soc. A: Math. Phys. Eng. Sci.*, *476*(2243), 20200110.

[5] Kothe, D., Lee, S., & Qualters, I. (2019). Exascale computing in the United States. *IEEE Comput. Sci. Eng.*, *21*(1), 17-29.

[6] Ltaief, H., Genton, M.G., Gratadour, D., Keyes, D.E., & Ravasi, M. (2022). Responsibly reckless matrix algorithms for HPC scientific applications. *Comput. Sci. Eng.*, *24*(4),12-22.

[7] Matsuoka, S., Domke, J., Wahib, M., Drozd, A., & Hoefler, T. (2023). Myths and legends in high-performance computing. *Int. J. High Perform. Comput. Appl.*, *37*(3-4), 245-259.

[8] McInnes, L.C., Heroux, M.A., Draeger, E.W., Siegel, A., Coghlan, S., & Antypas, K. (2021). How community software ecosystems can unlock the potential of exascale computing. *Nat. Comput. Sci.*, *1*, 92-94.

[9] Ootomo, H., Ozaki, K., & Yokota, R. (2023). DGEMM on integer matrix multiplication unit. Preprint, *arXiv:2306.11975*.

[10] Ootomo, H., & Yokota, R. (2022). Recovering single precision accuracy from tensor cores while surpassing the FP32 theoretical peak performance. *Int. J. High Perform. Comput. Appl.*, *36*(4), 475-491.

[11] Rosso, M., & Myers, A. (2021). A gentle introduction to GPU programming. *Better Scientific Software Blog*. Retrieved from https://bssw.io/blog_posts/a-gentle-introduction-to-gpu-programming.

[12] Trefethen, N. (2023, September 1). A bifurcation in Moore's law? *SIAM News*, *56*(7), p. 2.

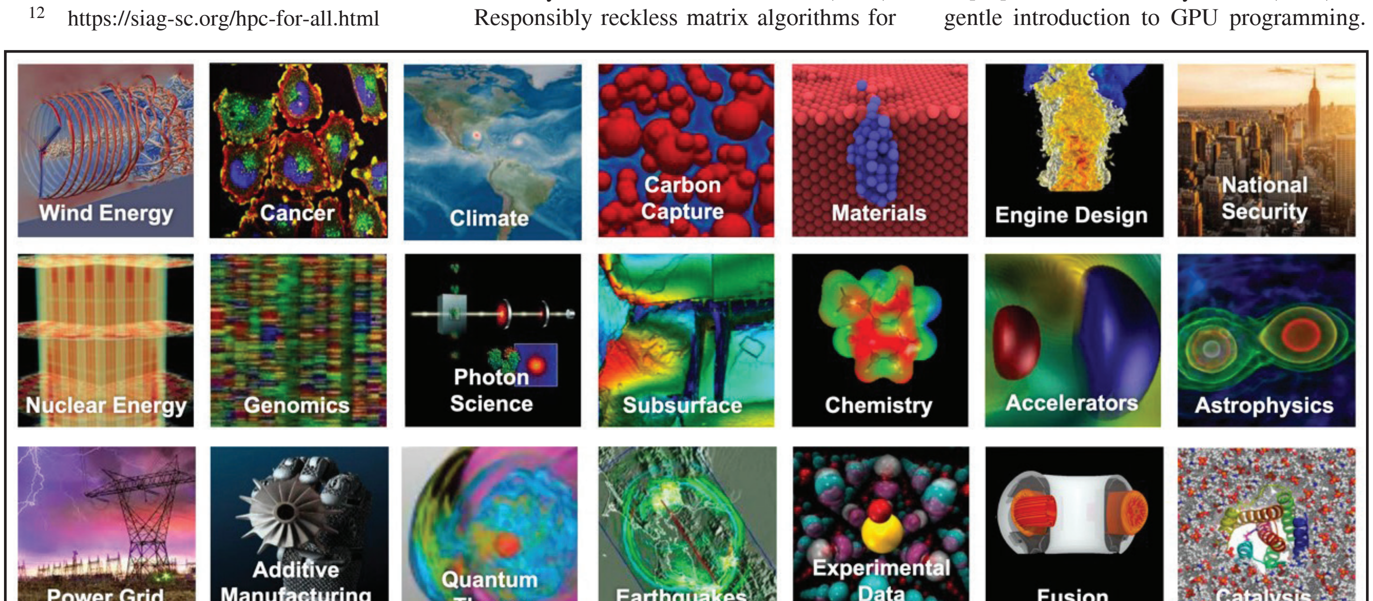


Figure 2. *Applications in scientific computing broadly impact science and society. For example, teams in the U.S. Department of Energy's Exascale Computing Project (ECP) are devising novel algorithms that are encapsulated in reusable software technologies to advance discovery in chemistry, materials, energy, Earth and space science, data analytics, optimization, artificial intelligence, and more [5]. Figure courtesy of ECP applications teams led by Andrew Siegel and Erik Draeger.*

Michael Bader is an associate professor in the TUM School of Computation, Information and Technology at the Technical University of Munich, where he works on hardware-aware algorithms and software for high-performance computing. He served as program director of the SIAM Activity Group on Supercomputing (SIAG/SC) from 2022-2023 and is co-chair of the 2024 SIAM Conference on Parallel Processing for Scientific Computing. Hatem Ltaief is a principal research scientist in the Extreme Computing Research Center at King Abdullah University of Science and Technology (KAUST). His research interests include parallel numerical algorithms and performance optimizations for manycore architectures. Ltaief served as vice chair of SIAG/SC from 2022-2023 and leads SIAG/SC outreach activities. Lois Curfman McInnes is a senior computational scientist and Argonne Distinguished Fellow in the Mathematics and Computer Science Division at Argonne National Laboratory. Her work focuses on high-performance scientific computing, with an emphasis on scalable numerical libraries and community collaboration towards productive and sustainable software ecosystems. She served as chair of SIAG/SC from 2022-2023. Rio Yokota is a professor in the Global Scientific Information and Computing Center at the Tokyo Institute of Technology. His research interests lie at the intersection of high-performance computing and machine learning. Yokota served as secretary of SIAG/SC from 2022-2023 and leads the SIAG/SC Committee, whose members help plan various SIAG/SC initiatives.

InsideSIAM
Conferences, books, journals, and activities of Society for Industrial and Applied Mathematics

SIAM | CONFERENCES

A Place to Network and Exchange Ideas

Upcoming Deadlines

SIAM Conference on Parallel Processing for Scientific Computing (PP24)

March 5–8, 2024 | Baltimore, Maryland, U.S.
go.siam.org/pp24 | #PP24

ORGANIZING COMMITTEE CO-CHAIRS
Michael Bader, *Technische Universität München, Germany*
Anshu Dubey, *Argonne National Laboratory, U.S.*

SIAM International Meshing Roundtable Workshop 2024 (SIAM IMR24)

March 5–8, 2024 | Baltimore, Maryland, U.S.
go.siam.org/imr24 | #SIAMIMR24

CONFERENCE CHAIR
Scott Mitchell, *Sandia National Laboratories, U.S.*

PP24 and IMR24 PRE-REGISTRATION DEADLINE
February 5, 2024: Disconnect time is 11:59 p.m. Eastern Time

PP24 and IMR24 HOTEL RESERVATION DEADLINE
February 5, 2024: Deadline is 5:00 p.m. local time

SIAM Conference on Nonlinear Waves and Coherent Structures (NWCS24)

June 24–27, 2024 | Baltimore, Maryland, U.S.
go.siam.org/nwcs24 | #SIAMNWCS24

ORGANIZING COMMITTEE CO-CHAIRS
Panayotis Kevrekidis, *University of Massachusetts, U.S.*
Anna Vainchtein, *University of Pittsburgh, U.S.*

SUBMISSION AND TRAVEL AWARD DEADLINES
December 27, 2024: Contributed Lecture, Poster, & Minisymposium Presentation Abstracts
March 25, 2024: Travel Fund Application Deadline

THE FOLLOWING THREE CONFERENCES ARE CO-LOCATED:

SIAM Annual Meeting (AN24)

July 8–12, 2024 | Spokane, Washington, U.S.
go.siam.org/an24 | #SIAMAN24

CONFERENCE CO-CHAIRS
Michael P. Friedlander, *University of British Columbia, Canada*
Anna Mazzucato, *Pennsylvania State University, U.S.*

SIAM Conference on Applied Mathematics Education (ED24)

July 8–9, 2024 | Spokane, Washington, U.S.
go.siam.org/ed24 | #SIAMED24

ORGANIZING COMMITTEE CO-CHAIRS
Mario Banuelos, *California State University, Fresno, U.S.*
Roummel Marcia, *University of California, Merced, U.S.*

SIAM Conference on Discrete Mathematics (DM24)

July 8–11, 2024 | Spokane, Washington, U.S.
go.siam.org/dm24 | #SIAMDM24

ORGANIZING COMMITTEE CO-CHAIRS
Peter Keevash, *University of Oxford, United Kingdom*
Blair D. Sullivan, *University of Utah, U.S*

AN24, ED24, and DM24 SUBMISSION AND TRAVEL AWARD DEADLINES
January 10, 2024: Minisymposium Proposal Submissions
January 31, 2024: Contributed Lecture, Poster, Miniposterium (AN24 only), and Minisymposium Presentation Abstracts
April 8, 2024: Travel Fund Application Deadline

Information is current as of November 16, 2023. Visit *siam.org/conferences* for the most up-to-date information.

Upcoming SIAM Events

ACM-SIAM Symposium on Discrete Algorithms
January 7–10, 2024
Alexandria, Virginia, U.S.
Sponsored by the SIAM Activity Group on Discrete Mathematics and the ACM Special Interest Group on Algorithms and Computation Theory

SIAM Symposium on Algorithm Engineering and Experiments
January 7–8, 2024
Alexandria, Virginia, U.S.

SIAM Symposium on Simplicity in Algorithms
January 8–10, 2024
Alexandria, Virginia, U.S.

SIAM Conference on Uncertainty Quantification
February 27–March 1, 2024
Trieste, Italy
Sponsored by the SIAM Activity Group on Uncertainty Quantification

SIAM Conference on Parallel Processing for Scientific Computing
March 5–8, 2024
Baltimore, Maryland, U.S.
Sponsored by the SIAM Activity Group on Supercomputing

SIAM International Meshing Roundtable Workshop
March 5–8, 2024
Baltimore, Maryland, U.S.

SIAM International Conference on Data Mining
April 18–20, 2024
Houston, Texas, U.S.
Sponsored by the SIAM Activity Group on Data Science

SIAM Conference on Applied Linear Algebra
May 13–17, 2024
Paris, France
Sponsored by the SIAM Activity Group on Linear Algebra

SIAM Conference on Mathematical Aspects of Materials Science
May 19–23, 2024
Pittsburgh, Pennsylvania, U.S.
Sponsored by the SIAM Activity Group on Mathematical Aspects of Materials Science

SIAM Conference on Imaging Science
May 28–31, 2024
Atlanta, Georgia, U.S.
Sponsored by the SIAM Activity Group on Imaging Science

SIAM Conference on the Life Sciences
June 10–13, 2024
Portland, Oregon, U.S.
Sponsored by the SIAM Activity Group on Life Sciences

SIAM Conference on Mathematics of Planet Earth
June 10–12, 2024
Portland, Oregon, U.S.
Sponsored by the SIAM Activity Group on Mathematics of Planet Earth

FOR MORE INFORMATION ON SIAM CONFERENCES: *siam.org/conferences*

SIAM | MEMBERSHIP

Network | Access | Outreach | Lead

Renew Your SIAM Membership

If you haven't already renewed your SIAM membership for 2024, go to ***my.siam.org*** to renew online or download a PDF of your renewal invoice. This will ensure uninterrupted access to your subscriptions to *SIAM News, SIAM Review*, and *Unwrapped*, and to your discounts on SIAM books, journals, and conferences. ***Free student members: Don't forget that you also need to renew to continue receiving benefits in 2024.***

Make the Most of Your Membership: Join an Activity Group!

SIAM Activity Groups (SIAGs) provide a focused forum for SIAM members interested in exploring specialized areas of applied mathematics and computational science. SIAGs organize conferences and minisymposia, disseminate newsletters, host online communities on SIAM Engage, and award prizes. SIAG members receive targeted communications from peers, access to electronic membership directories and community archives, and additional discounts on SIAG-sponsored conferences. ***Activity group membership is open only to SIAM members and dues are just $15 per year. Student members receive two free activity group memberships and Outreach members receive one free activity group membership.***

Algebraic Geometry (SIAG/AG)*
Chair: Greg Blekherman (01/01/22–12/31/23)
Websites: *siam.org/Activity-Groups/Algebraic-Geometry; wiki.siam.org/siag-ag*
Prize: SIAG/AG Early Career Prize

Analysis of Partial Differential Equations (SIAG/APDE)
Chair: Björn Sandstede (01/01/23–12/31/24)
Websites: *siam.org/Activity-Groups/APDE; siags.siam.org/siagapde*
Prizes: SIAG/APDE Best Paper Prize, SIAG/APDE Early Career Prize

Applied and Computational Discrete Algorithms (SIAG/ACDA)
Chair: Blair Sullivan (01/01/23–12/31/24)
Website: *siam.org/Activity-Groups/ACDA*
Prize: SIAG/ACDA Early Career Prize

Applied Mathematics Education (SIAG/ED)
Chair: Shelley Rohde Poole (01/01/23–12/31/24)
Upcoming conference: SIAM Conference on Applied Mathematics Education (ED24), July 8–9, 2024, Spokane, Washington (co-located with SIAM Annual Meeting)
Website: *siam.org/Activity-Groups/applied-mathematics-education*

Computational Science and Engineering (SIAG/CSE)
Chair: Judith Hill (01/01/23–12/31/24)
Websites: *siam.org/Activity-Groups/computational-science-and-engineering; https://siagcse.github.io/*
Prizes: SIAG/CSE Early Career Prize, SIAG/CSE Best Paper Prize

Control and Systems Theory (SIAG/CST)*
Chair: Jacquelien Scherpen (01/01/22–12/31/23)
Website: *siam.org/Activity-Groups/control-and-systems-theory*
Prizes: SIAG/CST Prize, SIAG/CST Best SICON Paper Prize

Data Science (SIAG/DATA)*
Chair: Ilse Ipsen (01/01/22–12/31/23)
Website: *siam.org/membership/activity-groups/detail/data-science*
Prizes: SIAG/DATA Career Prize, SIAG/DATA Early Career Prize

Discrete Mathematics (SIAG/DM)
Chair: Dana Randall (01/01/23–12/31/24)
Upcoming conference: ACM-SIAM Symposium on Discrete Algorithms (SODA24), January 7–10, 2024, Alexandria, Virginia
Website: *siam.org/Activity-Groups/discrete-mathematics*
Prize: Dénes König Prize

Dynamical Systems (SIAG/DS)*
Chair: Mason Porter (01/01/22–12/31/23)
Publication: Free subscription to SIAM Journal on Applied Dynamical Systems
Websites: *siam.org/Activity-Groups/DS; dsweb.siam.org*
Prizes: Jürgen Moser Lecture, J. D. Crawford Prize, Red Sock Award

NEW! Equity, Diversity, and Inclusion (SIAG/EDI)
Chair: Tamara G. Kolda (01/01/23–12/31/25)
Website: *www.siam.org/membership/activity-groups/equity-diversity-and-inclusion*

Financial Mathematics and Engineering (SIAG/FME)*
Chair: Igor Cialenco (01/01/22–12/31/23)
Websites: *http://wiki.siam.org/siag-fm/index.php/Main_Page*
Prizes: SIAG/FME Early Career Prize, SIAG/FME Conference Paper Prize

Geometric Design (SIAG/GD)
Chair: Hendrik Speleers (01/01/23–12/31/24)
Website: *siam.org/Activity-Groups/geometric-design; wiki.siam.org/siag-gd*
Prize: SIAG/GD Early Career Prize

Geosciences (SIAG/GS)
Chair: Masa Prodanovic (01/01/23–12/31/24)
Websites: *siam.org/Activity-Groups/geosciences; wiki.siam.org/siag-gs*
Prizes: SIAG/GS Career Prize, SIAG/GS Early Career Prize

Imaging Science (SIAG/IS)*
Chair: Hongkai Zhao (01/01/22–12/31/23)
Upcoming conference: SIAM Conference on Imaging Science (IS24), May 28–31, 2024, Atlanta, Georgia, U.S.
Websites: *siam.org/Activity-Groups/imaging-science; wiki.siam.org/siag-is*
Prizes: SIAG/IS Best Paper Prize, SIAG/IS Early Career Prize

Life Sciences (SIAG/LS)
Chair: Frederick Adler (01/01/23–12/31/24)
Publication: Free subscription to *SIAM Journal on Applied Dynamical Systems*
Website: *siam.org/Activity-Groups/life-sciences*
Prize: SIAG/LS Early Career Prize

Linear Algebra (SIAG/LA)
Chair: Melina A. Freitag (01/01/22–12/31/24)
Websites: *siam.org/Activity-Groups/linear-algebra; siags.siam.org/siagla*
Prizes: SIAG/LA Best Paper Prize, SIAG/LA Early Career Prize

Mathematical Aspects of Materials Science (SIAG/MS)
Chair: Irene Fonseca (01/01/23–12/31/24)
Websites: *siam.org/Activity-Groups/mathematical-aspects-of-materials-science; wiki.siam.org/siag-ms*

Mathematics of Planet Earth (SIAG/MPE)
Chair: Pierre Lermusiaux (01/01/23–12/31/24)
Websites: *siam.org/Activity-Groups/mathematics-of-planet-earth; http://mpe.dimacs.rutgers.edu/*
Prize: SIAG/MPE Prize, SIAG/MPE Early Career Prize

Nonlinear Waves and Coherent Structures (SIAG/NWCS)
Chair: Gino Biondini (01/01/23–12/31/24)
Website: *siam.org/Activity-Groups/nonlinear-waves-and-coherent-structures*
Prizes: Martin Kruskal Prize Lecture, T. Brooke Benjamin Prize in Nonlinear Waves

Optimization (SIAG/OPT)
Chair: Luis Nunes Vicente (01/01/23–12/31/25)
Publication: *SIAG/OPT News and Views*
Websites: *siam.org/Activity-Groups/optimization; wiki.siam.org/wiki/siag-op*
Prizes: SIAG/OPT Best Paper Prize, SIAG/OPT Early Career Prize, SIAG/OPT Test of Time Award

Orthogonal Polynomials and Special Functions (SIAG/OPSF)
Publication: OP-SF Net
Websites: *siam.org/Activity-Groups/orthogonal-polynomials-and-special-functions; math.nist.gov/opsf*
Prize: Gábor Szegö Prize

Supercomputing (SIAG/SC)*
Chair: Lois Curfman McInnes (01/01/22–12/31/23)
Website: *siam.org/Activity-Groups/supercomputing*
Prizes: SIAG/SC Career Prize, SIAG/SC Early Career Prize, SIAG/SC Best Paper Prize

Uncertainty Quantification (SIAG/UQ)
Chair: Amy Braverman (01/01/23–12/31/24)
Website: *siam.org/Activity-Groups/uncertainty-quantification*
Prize: SIAG/UQ Early Career Prize

* Officer election in progress

Join a SIAM Activity Group when you renew your SIAM membership for 2024!

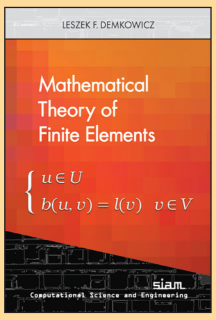
If you have already renewed, you can join an activity group (SIAG) online anytime at ***siam.org/Membership/Activity-Groups***. SIAG members receive many benefits, including being listed in an online SIAG-only membership directory and receiving a $15 registration discount for conferences sponsored by your activity group.

And don't forget—students can join two activity groups for free and outreach members can join one activity group for free!

FOR MORE INFORMATION ON SIAM MEMBERSHIP: ***siam.org/membership***

siam | BOOKS

Quality and Value in Mathematical Science Literature

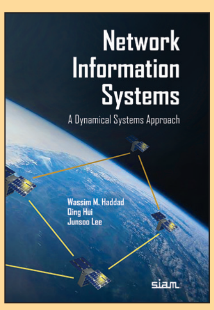


Mathematical Theory of Finite Elements

Leszek F. Demkowicz

This book discusses the foundations of the mathematical theory of finite element methods. The focus is on two subjects: the concept of discrete stability, and the theory of conforming elements forming the exact sequence. Both coercive and noncoercive problems are discussed.. Following the historical path of development, the author covers the Ritz and Galerkin methods to Mikhlin's theory, followed by the Lax–Milgram theorem and Cea's lemma to the Babuska theorem and Brezzi's theory. He finishes with an introduction to the discontinuous Petrov–Galerkin (DPG) method with optimal test functions. The book also includes a unique exposition of the concept of discrete stability and the means to guarantee it as well as a coherent presentation of finite elements forming the exact grad-curl-div sequence.

2023 / xi + 204 pages / Softcover / 978-1-61197-772-1
List $79.00 / SIAM Member $55.30 / CS28



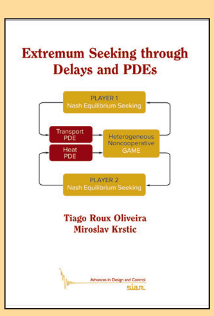
Network Information Systems: *A Dynamical Systems Approach*

Wassim M. Haddad, Qing Hui, and Junsoo Lee

This text presents a unique treatment of network control systems. Drawing from fundamental principles of dynamical systems theory and dynamical thermodynamics, the authors develop a continuous-time, discrete-time, and hybrid dynamical system and control framework for linear and nonlinear large-scale network systems. The proposed framework extends the concepts of energy, entropy, and temperature to undirected and directed information networks.

2023 / xiv + 622 pages / Hardcover / 978-1-61197-753-0
List $114.00 / SIAM Member $79.00 / OT191

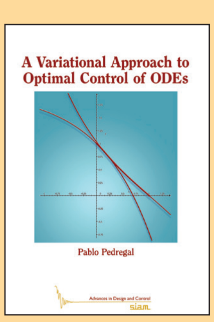
Extremum Seeking through Delays and PDEs



Tiago Roux Oliveira and Miroslav Krstic

This is the first book on this topic, and it expands the scope of applicability of the extremum seeking method, from static and finite-dimensional systems to infinite-dimensional systems. Readers will find numerous algorithms for model-free real-time optimization are developed and their convergence guaranteed. In addition, extensions from single-player optimization to noncooperative games, under delays and PDEs, are provided; the delays and PDEs are compensated in the control designs using the PDE backstepping approach; and stability is ensured using infinite-dimensional versions of averaging theory, and accessible and powerful tools for analysis.

2023 / xviii + 442 pages / Hardcover / 978-1-61197-734-9
List $119.00 / SIAM Member $83.30 / DC41



A Variational Approach to Optimal Control of ODEs

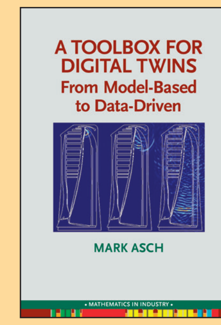
Pablo Pedregal

This self-contained book presents in a unified, systematic way the basic principles of optimal control governed by ODEs. Using a variational perspective, the author incorporates important restrictions like constraints for control and state, as well as the state system itself, into the equivalent variational reformulation of the problem. The fundamental issues of existence of optimal solutions, optimality conditions, and numerical approximation are then examined from this variational viewpoint. Inside, readers will find a unified approach to all the basic issues of optimal control, academic and real-world examples testing the book's variational approach, and a rigorous treatment stressing ideas and arguments.

2022 / xii + 189 pages / Softcover / 978-1-611977-10-3
List $69.00 / SIAM Member $48.30 / DC39

Upcoming Exhibitions

SIAM will be exhibiting at NeurIPS 2023 (December 10–16 in New Orleans) and CDC 2023 (December 13–15 in Singapore). If you're attending either, stop by and say hello!



A Toolbox for Digital Twins: *From Model-Based to Data-Driven*

Mark Asch

This book brings together the mathematical and numerical frameworks needed for developing digital twins. Starting from the basics—probability, statistics, numerical methods, optimization, and machine learning—and moving on to data assimilation, inverse problems, and Bayesian uncertainty quantification, the book provides a comprehensive toolbox for digital twins. Emphasis is also placed on the design process, denoted as the "inference cycle," the aim of which is to propose a global methodology for complex problems.

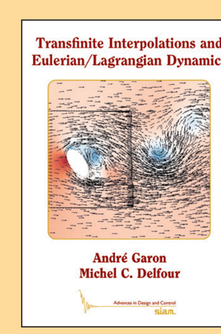
2022 / xxiv + 832 pages / Softcover / 978-1-611976-96-0
List $120.00 / SIAM Member $84.00 / MN06

Business Dynamics Models: *Optimization-Based One Step Ahead Optimal Control*

Eugenius Kaszkurewicz and Amit Bhaya

This book introduces optimal control methods, formulated as optimization problems, applied to business dynamics problems. It includes solutions that provide a rationale for the use of optimal control and guidelines for further investigation into more complex models, as well as formulations that can also be used in a so-called flight simulator mode to investigate different complex scenarios. The text offers a modern programming environment (Jupyter notebooks in JuMP/Julia) for modeling, simulation, and optimization, and Julia code and notebooks are provided on a website for readers to experiment with their own examples.

2022 / xxii + 190 pages / Softcover / 978-1-61197-730-1
List $89.00 / SIAM Member $62.30 / DC40



Transfinite Interpolations and Eulerian/Lagrangian Dynamics

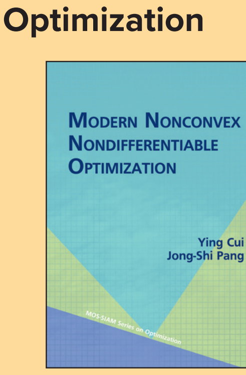
André Garon and Michel C. Delfour

This book introduces transfinite interpolation as a generalization of interpolation of data prescribed at a finite number of points to data prescribed on a geometrically structured set such as a piece of curve, surface, or submanifold. The time-invariant theory is readily extended to a moving/deforming data set whose dynamics is specified in a Eulerian or Lagrangian framework. The resulting innovative tools cover a very broad spectrum of applications in fluid mechanics, geometric optimization, and imaging.

2022 / xx + 268 pages / Hardcover / 978-1-611976-94-6
List $89.00 / SIAM Member $62.30 / DC38

Modern Nonconvex Nondifferentiable Optimization

Ying Cui and Jong-Shi Pang



Starting with the fundamentals of classical smooth optimization and building on established convex programming techniques, this research monograph provides a foundation and methodology for modern nonconvex nondifferentiable optimization by providing readers with theory, methods, and applications of nonconvex and nondifferentiable optimization in statistical estimation, operations research, machine learning, and decision making.

2021 / xviii + 774 pages / Hardcover / 978-1-611976-73-1
List $119.00 / SIAM Member $83.30 / MO29

To order, visit the SIAM bookstore: *bookstore.siam.org*

Or call toll-free in U.S. and Canada: 800-447-SIAM / worldwide: +1-215-382-9800

Do you live outside North or South America?

Order from Eurospan *eurospanbookstore.com/siam* for speedier service and free shipping.
Eurospan honors the SIAM member discount. Contact customer service (*service@siam.org*) for the code to use when ordering.

FOR MORE INFORMATION ON SIAM BOOKS: ***siam.org/books***

ATTENTION SIAM Math Modelers!

GOT A PROBLEM?

SIAM is Seeking Problem Ideas for Math Modeling Competition

MathWorks Math Modeling Challenge
A program of siam

$100,000 (£75,000) in SCHOLARSHIPS

What is M3 Challenge?

MathWorks Math Modeling (M3) Challenge is an internet-based, applied mathematics contest that takes place each year in February or March. High school juniors and seniors in the U.S. and sixth form students (age 16-19) in England and Wales may form and enter up to two teams of three to five students each per school. Teams are given 14 hours to solve an open-ended, applied math-modeling problem related to a real-world issue. Working collaboratively, students use math modeling to represent, analyze, make predictions and provide insight into current world issues. **Registration and participation are free**.

Past topics addressed issues such as substance abuse, food insecurity, climate change, car sharing, and modeling the cost, needs assessment, and placement of towers for maximizing access to the internet. View previous problem statements at ***M3Challenge.siam.org/resources/sample-problems***.

The goal of the Challenge is to motivate students to study and pursue careers in STEM disciplines, especially applied mathematics, computational science, data science, and technical computing. The problem is revealed to students only after they login on their selected Challenge day. Solutions are judged on the approach and methods used and the creativity displayed in problem solving and mathematical modeling. Extra credit in the form of technical computing scholarship awards is available for teams who opt to submit code.

Winners receive scholarship prizes totaling $100,000 (£75,000).

Problem structure

Within the problem statement, there should be three questions:

- Question One: A warm up — every serious team can answer.
- Question Two: The guts — framed so that every team can have some success and many teams will cover it well.
- Question Three: The discriminator — many teams will do something, while only a few will have striking results.
- Data — data that is provided or easily found is desirable to encourage students to use coding and technical computing in solution papers.

Honoraria

- $50 for problems found suitable to add to the M3 Challenge problem reserve "bank"
- $500 for problems selected from the reserve bank to be used as "the" Challenge problem

Required problem characteristics

- Accessibility to high school/sixth form students
- Suitability for solution in 14 hours
- Possibility for significant mathematical modeling
- Topic of current interest involving interdisciplinary problem solving and critical thinking skills
- Availability of enough **data** for a variety of approaches and depth of solutions (but no easily found answers)
- References identified that will be helpful for getting students started
- Submitted problem idea in the format of previous Challenge problems
- Potential to extend and enhance model using technical computing if a team chooses to do so.

Watch a video that explains M3 Challenge in one minute!
Go to YouTube and search on "About MathWorks Math Modeling Challenge"

Submit your ideas: *M3Challenge.siam.org/suggest-problems*

Contact SIAM for more information: M3Challenge@siam.org

M3Challenge.siam.org

MathWorks Math Modeling Challenge
A program of siam

A program of siam | Society for Industrial and Applied Mathematics

Sponsored by MathWorks

12/23