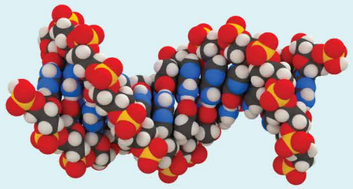
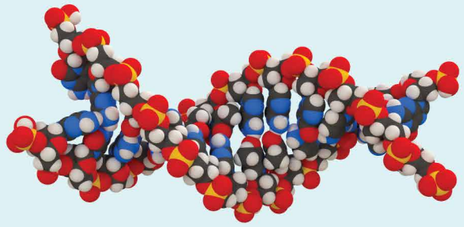
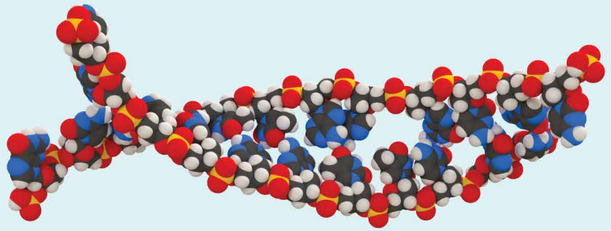
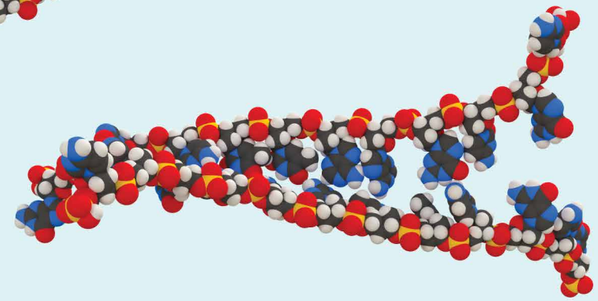
From Data to Information to Action

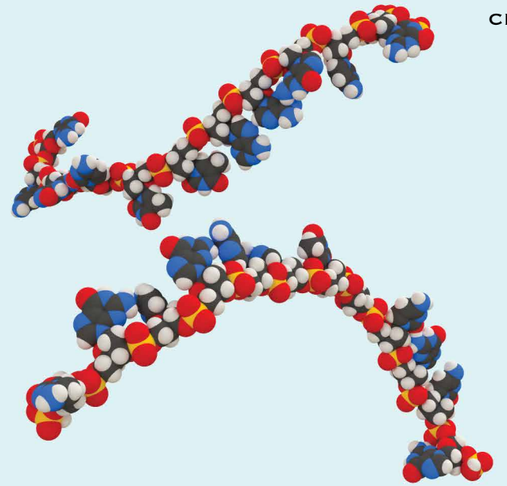
Computing + Mathematical Sciences Faculty











With an ever-increasing need to understand complex behaviors at the molecular level comes a high demand for computational methods. Professor Desbrun's group has proposed an efficient and scalable numerical method to run molecular dynamics simulations about 30 times faster than current algorithms. The new fine-grained parallelizable numerical scheme can handle a variety of simulations, from DNA strand unbinding (depicted here) and protein folding to nanotube resonators.

For millennia, engineers and applied scientists have brought mathematical tools to bear on problems impacting people, their lives, and their possessions. The Computing + Mathematical Sciences (CMS) faculty at Caltech are working in this tradition, creating tools and conducting research to move from data and problems to information and action. Their passion and research are rooted in the fundamentals and rigor of mathematics, with the ultimate goal of helping society make decisions and take action. Caltech students are heavily drawn to this approach, and to serve them better, the CMS faculty have created a new CMS PhD program.

ENGenious met with a subset of the CMS faculty to learn more about their interests and approach. The conversation explored the relationship between their research and energy, music, economics, special effects in movies, synthetic biology, and, of course, the nature of decision making.

"I provide tools for engineers," said Mathieu Desbrun, John W. and Herberta M. Miles Professor of Computing and Mathematical Sciences and the first Executive Officer of the CMS department. "So I'm no longer a bona fide engineer in the sense that I don't do big computations of tsunamis, but I do develop discretizations and computational methods so that other people can, including companies such as Schlumberger or Pixar."

Desbrun started in computer graphics before moving to the more theoretical field of applied geometry, doing so after an encounter with the late Caltech applied mathematician Jerrold Marsden, Carl F. Braun Professor of Engineering and Control and Dynamical Systems, who pointed out that some of his computer-graphics work on geometric discretization could be described by exterior calculus. "I had no idea what it was," says Desbrun. "But once he said this, I started scratching a little bit of the surface to see what he meant. And he was right!" It was a career changer: "I moved from graphics to becoming a tool designer for engineers, both in terms of computational methods and geometry processing."

The influence of this approach can be seen in the Information Science and Technology (IST) initiative, which was born out of the observation that information science on one side and science and engineering on the

> *"In the '60s when [computer graphics] started, there was no equipment, not even a monitor able to plot images . . . We have made a huge amount of progress; today's special effects in movies and video games are a visual testament to that. The impact on medical applications and parallel computing architecture is less visible but just as significant."*
>
> Mathieu Desbrun, *John W. and Herberta M. Miles Professor of Computing and Mathematical Sciences*

Mathieu Desbrun

other can lead to new synergies at their interface that give rise to whole new sets of insights in a variety of areas, including medicine, science, and society. "For example, quantum systems as systems which perform computation," says Peter Schröder, Shaler Arthur Hanisch Professor of Computer Science and Applied and Computational Mathematics. "What new insights for computation as well as physics does this allow? Or the insides of a cell as a giant network—like the Internet, with messages being sent everywhere. One can then bring in information theory (measures of information content and transmission bandwidth) to help understand regulatory networks in a cell."

Desbrun continues: "In the '60s when [computer graphics] started, there was no equipment, not even a monitor able to plot images. It was super complicated to do, but now graphics have been so successful that everybody has a graphics card with power that, back in the '60s, would have required a whole city full of computers. We have made a huge amount of progress; today's special effects in movies and video games are a visual testament to that. The impact on medical applications and parallel computing architecture is less visible but just as significant."

One of Desbrun's applied mathematics colleagues is Yizhao Thomas Hou, Charles Lee Powell Professor of Applied and Computational Mathematics, who is an expert in the very traditional research task of pulling patterns out of masses of incoming data, particularly the behavior of fluids, and has learned to do this using extremely fine-scale mathematical modeling.

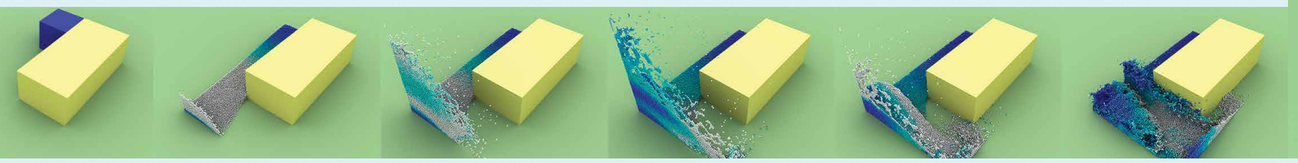
His research has expanded the scope of classic works like the earlier Euler and subsequent Navier-Stokes equations, which govern the motion of inviscid and viscous flows and are used in efforts to predict phenomena ranging from ocean currents to blood flow to weather. But the equations have run up against limits in attempts to expand them to wider parameters and smaller scales. A very well-known Millennium Problem is whether the solution of the Navier-Stokes equations will remain smooth for all time if one starts with sufficiently smooth initial data, or whether it will break down in finite time. A $1 million prize awaits the researcher who can answer this question, a prize Hou is seeking. While conducting this search, Hou and his colleagues recently discovered a scenario that leads to a previously unsuspected "singularity," an irregular point interrupting or redirecting flow, which provides a promising scenario for further investigation of the potential singularity.

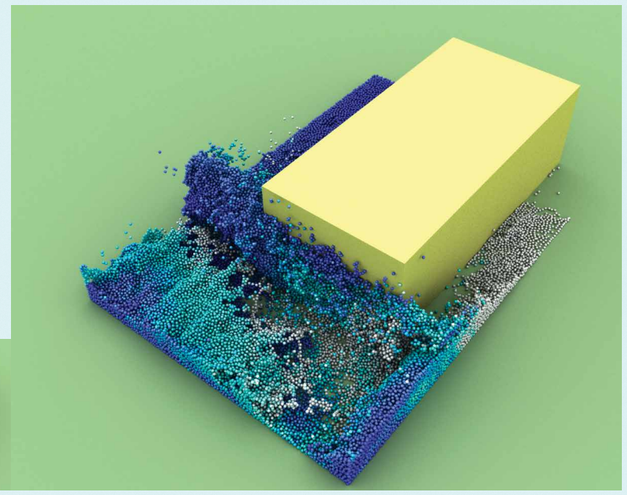
Some of Hou's early work in the area of fluid behavior modeling has had applications in the energy sector, where oil company engineers use it to simulate two-phase flow to enhance

Yizhao Thomas Hou

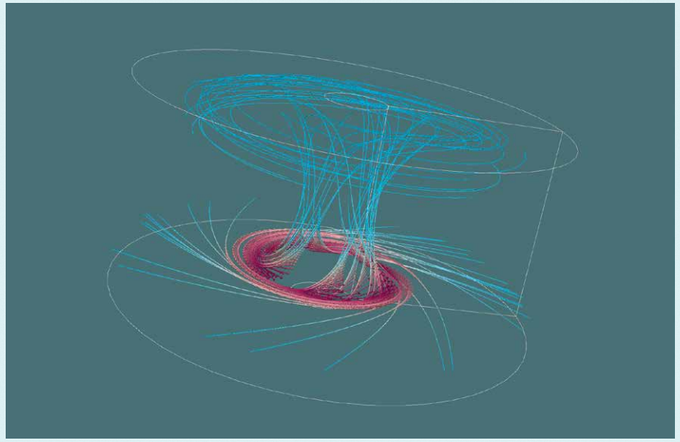
"You could use it on your cell phone to measure your pulse. It solves the optimization problem on the spot and sends the data to your doctor. The doctor can then determine if you really have a problem and are at risk or not."

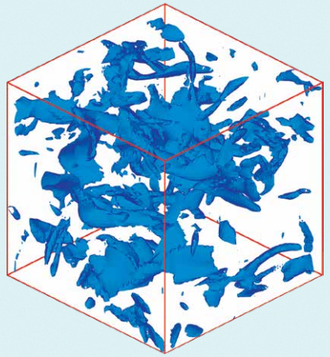
Yizhao Thomas Hou, *Charles Lee Powell Professor of Applied and Computational Mathematics*





Fluid motion is a visually rich and complex phenomenon that remains a challenge to reproduce numerically. While Lagrangian methods excel at generating fluid motion with few degrees of freedom, they often suffer from numerical artifacts that severely impact the liveliness of the flow. The Applied Geometry Lab recently proposed a radically different Lagrangian method using a particle-based approach to simulate incompressible fluid. Particles are considered as non-overlapping fluid parcels that partition the space occupied by the fluid through a time-evolving power diagram. By leveraging computational tools for power diagrams, the researchers formulated a time integrator for the "power" particles that precisely controls particle density and pressure forces, without kernel estimates or significant artificial viscosity. The versatility of the Applied Geometry Lab's solver (colors of particles indicate their velocity) is demonstrated in the "wall-confined dam break" sequence, simulated with 65,000 particles. It shows water first splashing against the opposite wall before ricocheting onto the inner wall (yellow box), then splashing at the end of the domain and finally settling.





Above: The velocity field and stream function plot of a tornado-like flow for the three-dimensional axisymmetric Navier-Stokes equations.

Left: Iso-surface plot of vorticity in a three-dimensional incompressible turbulent flow.

oil recovery. Hou has also found a new way to customize general analytic methods to specific formations. In addition, his theoretical efforts have found application in an area seemingly remote from oil recovery: in blood flow, specifically a mobile-device–based application that can sense, read, and analyze live data. "You could use it on your cell phone to measure your pulse," he says. "It solves the optimization problem on the spot and sends the data to your doctor. The doctor can then determine if you really have a problem and are at risk or not."

The dynamics of large structures, such as bridges, can also be diagnosed using Hou's work. "In the past, if we wanted to measure the forces on the bridge on a windy day as a truck was passing, it would have required a lot of time and expense," he explains. "But today, remote sensors on the bridge detect the frequency of vibrations." Theoretically, analysis of these data could be used to determine the strength of the forces. But the analysis requires sophisticated new mathematics, and after years of collaboration, Hou's team is closing in on the solution.

Professor Schröder is another creator of mathematical tools. "I don't build cars," he says. "I build engines. I build the motor underneath the hood that makes the machine purr. So what that means, practically speaking, is that I think about algorithms and numerical techniques that can take the physical laws that describe, for example, how a piece of cloth dangles in the wind, and then turn those physical laws into efficient computations so that the simulation can be used to move the shirt of a character in a Pixar movie."

The standard is high, he says. "The eye, because of our species's years of genetic optimization, is extremely good at being able to see whether something is real or not. To use an example from real life, you might see somebody walking down the street at a great distance where your eye can't actually tell their face, but you recognize the person by their walk. This is an example of how incredibly in tune we are to qualitative things. So in computer graphics for entertainment purposes, the measure of fidelity is to capture this in numerical ways. This is not all we do, but this is an important part of what we do. And here, as in other places, we have learned that you have to get the physics right."

Schröder loves the complications involved not just in getting it right but in doing so efficiently and elegantly. The ideals are algorithms that help this adaptive process: "algorithms that very quickly give us a rough idea. Then the same algorithm should be able to give us more and more precise answers as we give it more time."

Schröder's road to Caltech was unusual. "I left high school in Germany to travel around the world, and then to study psychology," he explains. After years of exploration, he trained as a shiatsu specialist and worked with clients in a private practice in Manhattan. Then a friend showed him the 1982 American science fiction film *Tron*, which was transformative. It led him to take a course on mathematics for computer graphics at a graphics conference in 1984. One of the lecturers in that course was a Caltech professor of computer science, Alan Barr. Little did Schröder know that he would be Professor Barr's colleague one day. Schröder's newly discovered passion for mathematics subsequently led him to the MIT Media Lab, a Princeton PhD, and then his faculty position in the CMS department.

He loves the CMS culture, which he says is about "bridge building, and not just bringing a technique from this field to that field but really having a new synergy occurring where both sides go, 'Wow, we can do all kinds of new things we didn't know how to do before.'" The CMS students, too, impress Schröder: "They have something burning inside of them like a fire that cannot be quenched."

Natural phenomena like cacti (left) exhibit characteristic branching patterns in an effort to maintain evenly spaced features. Schröder's algorithm allows one to quickly and automatically synthesize similar patterns that seamlessly cover arbitrary surfaces (right).

Peter Schröder

"I think about algorithms and numerical techniques that can take the physical laws that describe, for example, how a piece of cloth dangles in the wind, and then turn those physical laws into efficient computations so that the simulation can be used to move the shirt of a character in a Pixar movie."

Peter Schröder, *Shaler Arthur Hanisch Professor of Computer Science and Applied and Computational Mathematics; EAS Division Deputy Chair*

> *As my field of optimization moves forward, it aids decision making, turning into a standard, mature, and reliable tool that can be used easily and seamlessly to quickly obtain actionable and interpretable information from data.*

Venkat Chandrasekaran, *Assistant Professor of Computing and Mathematical Sciences and Electrical Engineering*

Venkat Chandrasekaran

The experience of Venkat Chandrasekaran, Assistant Professor of Computing and Mathematical Sciences and Electrical Engineering, is similar. "Much of the research that happens in CMS is grounded in the mathematical foundations, more so than at other computer science departments," he says. "And that comes across in the way we interact with students and is successfully transferred to them."

In his research, Chandrasekaran investigates the conceptual foundations of optimization, a branch of applied mathematics that focuses on designing the most efficient approach to accomplishing a task. This methodology is useful for engineering optimal machines and systems.

"People who work in my area have had impact in domains as varied as computational finance, medical imaging, aircraft design, and power flow in the smart grid," he says. "Statistics is an application domain that serves as a major motivation for a lot of my work in optimization."

CMS is an ideal place to carry out this work. "Within the Caltech community, it's special in that it's a very outward-looking department," Chandrasekaran explains. "I don't know of any other major research institution that has applied mathematicians and computer scientists and control and dynamical systems engineers sitting in the same internal department."

And the CMS effort is gathering momentum, he notes: "To borrow from the mission statement from the CMS graduate program, we are trying to do research in the pipeline that takes us from data to information to action. This last leg, going from information to action, is something that I think is unique to us. As my field of optimization moves forward, it aids decision making, turning into a standard, mature, and reliable tool that can be used easily and seamlessly to quickly obtain actionable and interpretable information from data."

Chris Umans

Chris Umans, Professor of Computer Science, works in closely adjoining areas using an approach he calls "understanding computation as a phenomenon." He hopes to build up "a framework in which we can think about computation and do things computationally in a principled way that's not hacking."

This means getting down to roots. "I spend a lot of time thinking about fundamental algorithms for fundamental problems," he says. "These are problems that people identify as fundamental because they're at the core of a lot of different applications. If we can improve performance on these, by finding more sophisticated math, then we can solve those problems in either a faster or in a fundamentally different way, which can affect many applications that build out from there."

Umans reflects on the importance of stepping back from the conventional approach and trying alternatives—not necessarily from computer science. "Computer science is a really young field, and mathematics has been around for thousands of years," he says. "The thing that seems to keep happening is that the kinds of questions that we as computer scientists ask are close to the kinds of questions that mathematicians are interested in, but not quite. So we get inspiration from the way that they've dealt with things."

The work of Leonard Schulman, Professor of Computer Science, has been inspired by—and amplifies in new directions—the work of Claude Shannon, who seven decades ago cre-

> *I spend a lot of time thinking about fundamental algorithms for fundamental problems. These are problems that people identify as fundamental because they're at the core of a lot of different applications.*

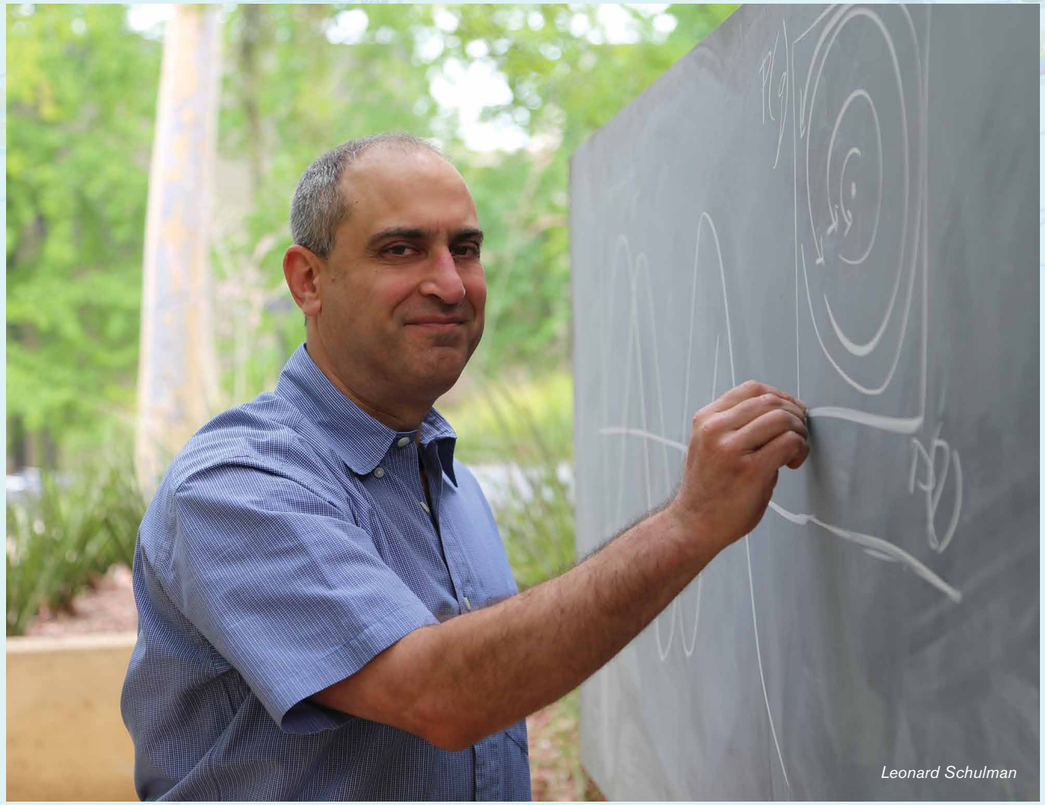
Chris Umans, *Professor of Computer Science*

ated the mathematical definition of information, including the possibility of error-free transmission despite any amount of noise. This is now a part of our everyday life. "Though no one using a cell phone really cares how the coding theory is done, it wouldn't be there if it weren't for coding theory," says Schulman.

Schulman is dealing with "the next generation of this problem, when communications are highly interactive and at very high rates of interaction." In these situations, he says, "with a huge amount of interaction happening and very short bursts of communication back and forth over a long period, orthodox Shannon error-coded communication can become unwieldy."

The consequences of a glitch in such a dense mix can be large. Even in human conversations, he notes, "a small misinterpretation can derail into a huge miscommunication or misunderstanding between the people and/or groups involved. This can happen with non-human communications, as well, and can be even worse when we get to the world of arbitrary network protocol. But if we try to solve it the old-fashioned way, by just putting a large amount of redundancy and error correction into each individual message, it would slow it down a significant amount—essentially an unbounded amount, unless an algorithm is found to speed it up."

Such communication conflicts, statistics, and algorithmic workarounds are basic parts of the Schulman research agenda. Running deeper is a motivation to maximize human communication: "There's been a lot of work in the field of algorithms and machine learning on [such questions as]: How do we analyze data? How do we cluster it? How do we find structure in it? But when you actually look at how scientists, business people, and government officials really use data, it is to make a decision. Therefore, what we really want to answer are not the academic questions of the existence of correlations but the more crucial one of analyzing whether correlations are causal or accidental."



Leonard Schulman

> *"When you actually look at how scientists, business people, and government officials really use data, it is to make a decision. Therefore, what we really want to answer are not the academic questions of the existence of correlations, but the more crucial one of analyzing whether correlations are causal or accidental."*
>
> Leonard Schulman, *Professor of Computer Science*

This is a very hard problem, but Schulman believes his group's recent work on a novel analytic model to distinguish between the two is promising, with potentially extremely far-reaching impact. More specifically, he explains, "the 'usual' way we determine causality is by running controlled experiments, such as giving half the subjects treatment A, half treatment B, and observing how they fare. But often we can't do that for a variety of reasons, including it being unethical or even impossible. So, like it or not, we have to extract inference about causality from purely observational data. In the most general framework, this is impossible, but with some extra conditions on the structure of the system, remarkably, it is sometimes possible—emphasis on 'sometimes.' Currently, the theory establishes this to be possible only under very restrictive conditions. My work is geared toward relaxing those assumptions. The ultimate goal is to be able to make scientifically rigorous statements about causal connections, based only on passive observational data. This could apply to many scenarios—in medicine, public health, educational policy, welfare policy, ecology, and the environment."

Assistant Professor of Computing and Mathematical Sciences Thomas Vidick's research aims to guide the problem-solving process at an earlier stage. He is working on mathematical methods to evaluate solutions to hard problems for which exact solutions are impossible to find—but approximate values are. "So it has to do with understanding what kind of problems are hard, and how hard," he says. "Traditional complexity theory has been able to say that finding the exact best solution to a certain problem can be computationally extremely hard. But what if you're happy with something that's 99% as good as the optimum—how hard is that?"

Vidick is working at the boundary of computer science and quantum information theory, with particular attention to computational possibilities and limits. His aim is to create tools that ideally can tell an engineer that "if this is the kind of question you're asking, then there's just no possible way to get an answer." He hopes his work would help guide researchers in determining in timely fashion that they can't go down a road, and that they must change their direction.

Vidick's work in cryptography considers the challenges of developing cryptosystems—the mechanisms used to transmit sensitive information, such as a credit card number, through public channels, including the Internet, that are based on the laws of quantum mechanics. Such cryptosystems can in theory be much more secure than classical cryptosystems. But quantum hackers have shown they are also much more prone to "side-channel attacks" that exploit vulnerabilities in the implementations. "My work is trying to develop cryptosystems that use quantum mechanics but do not need

Thomas Vidick

> *Introducing the weirdness of quantum mechanics . . . into the conceptual frameworks of complexity theory and cryptography produces insight . . .This is what makes the research challenging and exciting: you take a rich framework, you throw in a completely new ingredient, and you get beautiful chemistry!*

Thomas Vidick, *Assistant Professor of Computing and Mathematical Sciences*

to rely on the trustworthiness of the quantum devices used to implement the protocol," Vidick explains. "So that even if the devices malfunction, or the attacker has some control over them, the users will be able to detect this and abort the protocol."

This work fuels Vidick. "Introducing the weirdness of quantum mechanics, such as quantum entanglement, into the conceptual frameworks of complexity theory and cryptography produces insight into quantum mechanics, the global nature of entanglement, and properties such as the monogamy of quantum correlations," he says. "This is what makes the research challenging and exciting: you take a rich framework, you throw in a completely new ingredient, and you get beautiful chemistry!"

Katrina Ligett, who holds a joint appointment in computer science and economics with interests that bring

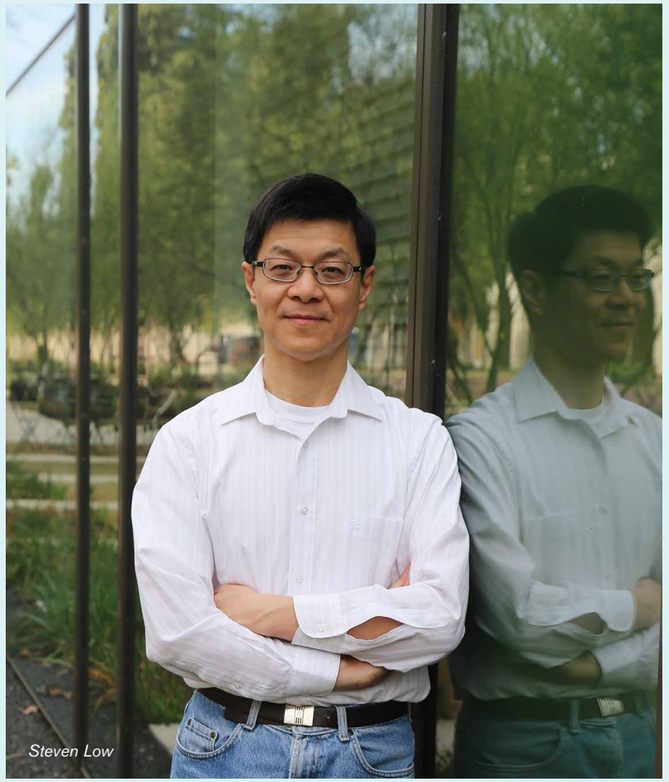
Katrina Ligett

together the Division of the Humanities and Social Sciences and the Division of Engineering and Applied Science, is also interested in data security and privacy. One of her research goals is giving formal guarantees on what a computation cannot leak. This fans out through a huge range of potential societal impacts and foci related to a fundamental question: What data privacy environment do we want?

"What's interesting about privacy is not so much what people or organizations are doing or not doing but rather the description of a data-leaky environment and strategies for dealing with it," Ligett says. "What could and should organizations and people do? Answering this question opens doors for investigation." She adds, "Key elements that need to be understood include the benefits of using this information, risks and costs of such use, the ways in which these

> *What's interesting about privacy is not so much what people or organizations are doing or not doing but rather the description of a data-leaky environment and strategies for dealing with it . . . There are lots of questions to be asked in this space, and I think it's a fun research place to play in.*

Katrina Ligett, *Assistant Professor of Computer Science and Economics*


Steven Low

> *Our research . . . starts by assuming that there will be a lot of renewables and we are going to have a lot of active endpoints that are intelligent but yet doing their own things. Then we ask, 'What are the new fundamental challenges that will arise?' These challenges are not only in engineering but also in economics. How do we design markets to incentivize the right behavior?*
>
> Steven Low, *Professor of Computer Science and Electrical Engineering*

data are transacted on, bought, sold, computed on, and tracked."

She goes on to explain that "almost everyone has to make these risk-benefit calculations, in a legal and social environment. Society has to decide the rules governing them. So I'm interested in really starting afresh in how we think about all of these interactions that we have with personal information and trying to figure out if we can do it differently. There are lots of questions to be asked in this space, and I think it's a fun research place to play in."

The work of Steven Low, Professor of Computer Science and Electrical Engineering, is deeply involved with information and energy infrastructure, which he says have "completely changed the way we live and work since their overlapping inceptions about 130 years ago."

While information infrastructure led to the Internet in the past few decades, the power network has not undergone much of a transformation. "California Edison still has transformers which are many decades old and working fine," says Low. "But the power network is now undergoing a drastic change on the order of magnitude of the information revolution that produced billion-dollar players like Google and Facebook."

Low is studying how this new transformation will proceed, and his efforts include calculations to determine how to optimize the process. "By changing the characteristics of the power grid, we can control it," he says. "We can minimize the loss. We can route power to avoid congestion on the grid. We can better guarantee power quality at the extreme loads. Power electronics are very important pieces of the equipment that will allow us to change the power grid and avoid congestion in ways that were not possible before."

The new vision of the grid is made up of "a network of intelligent endpoints that allow us to control much more actively, to close loops much faster, and to improve the efficiency, robustness, reliability, and security of the entire system in a way that we cannot do today," Low explains. "Our research at Caltech on the smart grid starts by assuming that there will be a lot of renewables and we are going to have a lot of active endpoints that are intelligent but yet doing their own things. Then we ask, 'What are the new fundamental challenges that will arise? These challenges are not only in engineering but also in economics. How do we design markets to incentivize the right behavior?"

It's a complicated undertaking. "To solve this huge energy problem, we not only need power system expertise but also control and dynamical systems, computer science, applied math, and economics," Low says.

Richard Murray

> *We don't know yet whether or not a decade or two from now, synthetic biology is something that we all take for granted and we go into our house and it's got a whole bunch of biological components that react to us being there and do smart things.*
>
> Richard Murray, *Thomas E. and Doris Everhart Professor of Control and Dynamical Systems and Bioengineering*

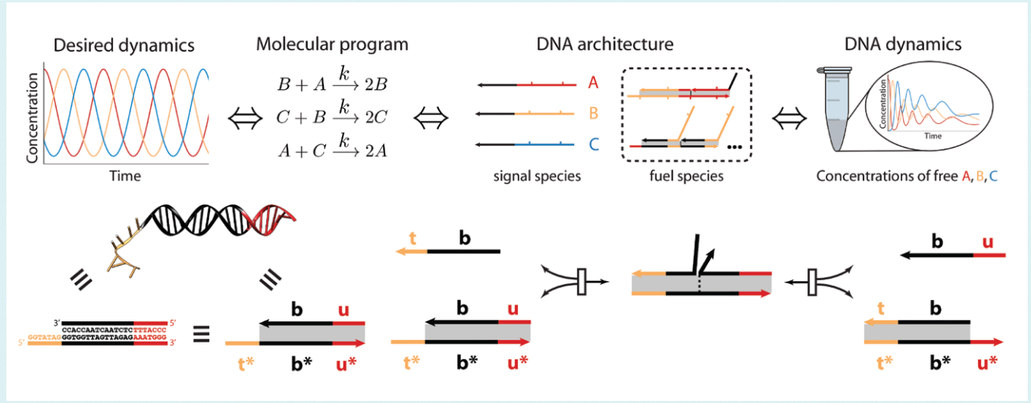
> *We've built a DNA system that oscillates. So it's like a little biochemical clock, but no enzymes are involved . . . It's a dynamical system, and computer scientists love systems that have programmable behaviors.*
>
> Erik Winfree, *Professor of Computer Science, Computation and Neural Systems, and Bioengineering*

"This is why CMS is the ideal place. We look at the underlying fundamental core, especially the mathematical aspects of those problems, and bring them together."

Mathematical tools to regulate networks and systems are also one of the keys to the work of Richard Murray, Thomas E. and Doris Everhart Professor of Control and Dynamical Systems and Bioengineering. Murray's current focus is another growing field, biological systems—specifically biomolecular feedback systems.

Application of biology network technology is still far from the explosive everyday application level of the data networks that concern Low. So discussing the societal impacts is difficult, says Murray, "because we're not there yet. I think what people would like to be able to do is build systems out of biological parts, DNA, RNA, proteins, in ways that perform useful functions. It can mean environmental remediation. It can mean just useful devices that process information and remember things and compute. We're at a very early stage. We don't know yet whether or not a decade or two from now, synthetic biology is something that we all take for granted and we go into our house and it's got a whole bunch of biological components that react to us being there and do smart things."

Erik Winfree



But Murray is confident that if the fundamental research moves to global change, Caltech will be part of it. His progression to studying biological systems was evolutionary. "I started in mechanical engineering because I was interested in robotics, but my degrees are in electrical engineering, so from the beginning I was bringing disciplines together," he says. "Then I began working with people in computer science and got interested in the role of feedback and control theory in biological systems. I wanted to explore the potential role of my field of control theory in biological, biochemical, and biomolecular applications—which led to what is now probably two-thirds of my group focusing on synthetic biology and the other third continuing to do things related to more traditional electromechanical systems and the software that sits on top of it."

Murray maintains that Caltech offers an ideal environment for movement in new research directions. "I decided to research synthetic biology," he says, and "that meant I needed a relatively large wet lab. But I didn't have to force Caltech to do something that it didn't want to do, but rather share what I was excited about and the specifics of what I needed to be successful. Then a way was found to make all of it work. This type of flexibility, involving helping the faculty move in new directions by providing seed funds, is one of the special things about Caltech and a key to our success."

Erik Winfree, Professor of Computer Science, Computation and Neural Systems, and Bioengineering, is also studying information processing in biomolecular systems, a research destination he arrived at via an unexpected route. As a high-school student, Winfree loved mathematics but hated "wet science"—"I decided I would never do biology or chemistry," he says. But now he, like Murray, has a bioengineering wet lab where he is adapting and programming molecules

This figure shows the framework for a molecular programming compiler based on strand displacement. In the first step, an abstract chemical reaction network is designed to obtain the desired chemical kinetics; in the second step, this molecular program is translated to a specification for DNA "signals" and DNA "fuels" that will carry out the desired reactions; and in the third step, specific DNA sequences are designed and synthesized and observed in the laboratory. Interactions between the DNA signals and fuels take place in the absence of enzyme or other biological components, making use of a mechanism known as toehold-mediated strand displacement, in which an incoming single-stranded DNA molecule can bind to a toehold (yellow and red) and displace one strand of a helix that contains an identical subsequence (black).

Adam Wierman

> *Caltech as a whole is going to benefit from investing in developing an understanding of how to process large problems, and how to store and operate on large datasets. Despite our small size in CMS, almost everything at Caltech is touched in some part by computing and the CMS faculty.*

Adam Wierman, *Professor of Computer Science; Executive Officer for Computing and Mathematical Sciences*

to carry and transmit information and use it to control chemical system behavior. The work centers on the natural information carriers DNA and RNA, but the scope is much wider.

In fact, he says, "we now have a general-purpose way of building molecular machines that implement arbitrary chemical reaction networks. So although we haven't done it for every possible chemical reaction network, we have a very detailed engineering argument that if you give us a set of reactions that does something interesting—$A + B$ goes to $X + Y$, Z goes to $P + Q + R$, etc.—we can hit 'go' and it will get compiled down into DNA molecules. Although we are now studying dynamical behavior, the original goal was to establish a connection between crystallization and self-assembly on the one side and computer science and algorithms on the other side." Hitting "go" is different for biochemical and electronic computing; translating electronic circuits into chemical ones will not be achieved immediately. But Winfree is quick to note recent advances: "We've built a DNA system that oscillates. So it's like a little biochemical clock, but no enzymes are involved. No fancy chemistry. Just a basic machinery that releases DNA strands, changing who's partnering with who. It's a dynamical system, and computer scientists love systems that have programmable behaviors."

Researchers with a range of skills and specialties are working on getting to "go." "A lot of us are mathematicians, with a very strong emphasis on the theoretical, rigorous foundations of what we do," says Winfree. "There's a metaphor of the cathedral of science, where every scientist has the opportunity to put one brick on the cathedral. We want a strong building, and if we build with a bad brick that falls apart and cracks, well, that's not so good. Therefore, emphasizing the mathematical foundations is important for creating a really solid understanding that other people can build on."

Such rigor is also central to the research approach of CMS professor Adam Wierman, who is part of the Linde Institute of Economic and Management Sciences and the Resnick Sustainability Institute. His basic research thrust concerns the care, management, and security of networked systems, including those distributing data and power. The engineering issues involved in these areas relate directly to economic and social inputs and consequences, as the Linde Institute connection emphasizes. "There's this real interplay between engineering and economics. We try to do a lot of foundational theory with respect to that interaction across different disciplines," Wierman says.

One economic area central to his work is IT energy demands, which, according to Wierman, are huge and growing. The groundwork has changed. "Memory is no longer the bottleneck. It's very cheap. Data is the bottleneck," he explains. "And access to these terabytes of cheap information raises energy issues. If access to all contents of thousands of servers has to always be available instantly, the power demand is maximized: everything has to be working at top speed all the time. But if the service can be tailored to urgency, large energy savings are possible. Even more is possible by two-way communication between the power grid and data system controllers."

The skills necessary for such technical and economic balancing acts have been emerging from the work of several CMS faculty and have been one of the inspirations for the new CMS degree program. "We have tried to develop a new intellectual core and wrap a PhD program around that rather than forcing people to merge their silos and go outside of their traditional fields to add to a degree what they actually need and care about," says Wierman. "We think this approach is core to doing science, engineering, and information sciences in the next 10 years, an area that includes optimization, machine learning, stochastic processes, algorithms, networks, and economics. This is how the core curriculum in the new PhD program came together, and we are excited to welcome the first class this fall."

To Wierman, this is a key movement in the right direction. "Data is the bottleneck not only to development of applications, programs, and research but to science and engineering progress in general," he says. "Caltech as a whole is going to benefit from investing in developing an understanding of how to process large problems, and how to store and operate on large datasets. Despite our small size in CMS, almost everything at Caltech is touched in some part by computing and the CMS faculty."

Joel Tropp

Another architect of the new CMS PhD program is Joel Tropp, Professor of Applied and Computational Mathematics. Tropp works in the field of parsimonious modeling, also called sparse approximation. An observer in imaging science, machine learning, communications, or statistics often tries to analyze a flow of data to find patterns, assuming the data are the result of an undetermined but determinable mathematical relationship. This is a difficult general problem, but Tropp has found algorithms that help find such mathematical ties in specific cases.

He uses music as an analogy to illustrate, starting from the point of view of a mechanical listener: "It turns out that if recordings didn't have any structure, they would sound like static, whereas they tend to have dominant frequency components, much stronger tones and overtones. And they're also localized in time and space, so there are silences."

He continues: "Sheet music is a very efficient way to represent what can be a very complicated piece of music. Thus the idea is that if we can identify this kind of representation for data, then we can compress the data significantly. This is the key. Once we realize that there's an underlying pattern, then we can write down the piece of music much more efficiently."

> *Sheet music is a very efficient way to represent what can be a very complicated piece of music. Thus the idea is that if we can identify this kind of representation for data, then we can compress the data significantly.*

Joel Tropp, *Professor of Applied and Computational Mathematics*

> *Machine learning can be used to help build smarter cancer detection methods using imaging analysis tools. It takes a radiologist's time to understand an X-ray, and researchers have been thinking about using more automated techniques for imaging analysis to improve the detection process both in time and accuracy.*

Yisong Yue, *Assistant Professor of Computing and Mathematical Sciences*

Yisong Yue

Coincidentally, Tropp's research has found applications in sound analysis, where observers are trying to pick unknown signals out of a flow. Using the right software to compress a representation of the signal, he has found ways to improve the analysis and make finding the signal easier. A similar result comes from tables of information, referred to as randomized linear algebra. "When we are trying to find structure in a very big matrix or a table of data, surprisingly, we can identify the structure automatically just by taking random combinations of the data that we've seen," says Tropp. "The random combinations contain the same underlying structure as the whole, which the algorithm more efficiently finds."

Yisong Yue, Assistant Professor of Computing and Mathematical Sciences, also works on ways to understand masses of data in a less abstract context. He studies machine learning, "the automated process of turning data and experience into knowledge and actionable items. Today, when we do anything on the Internet that is commercial, there's some sort of machinery under the hood that's trying to predict what it is we are interested in. The predictions we see can be helpful because they could help shorten the amount of time it takes us to find what we are look-

ing for. These are machine-learning algorithms that take the history of other people's purchases and browsing behavior and predict what we might be interested in."

This growing field has come a long way in recent times. "If you go back 10 years and try to shop online, you would think it is a disaster how slow and inefficient it was," says Yue. Since then, machine learning has stepped in "to convert massive amounts of data, in the form of logs of what people have done in the past, to make things more efficient."

Shopping is not the only process that machine learning can improve. "Machine learning can be used to help build smarter cancer detection methods using imaging analysis tools," Yue explains. "It takes a radiologist's time to understand an X-ray, and researchers have been thinking about using more automated techniques for imaging analysis to improve the detection process both in time and accuracy."

Another area of interest for Yue is video and tracking data. "Now there is huge interest in studying how different genes impact the brain," he says. "Neuroscientists manipulate the genes of test animals such as fruit fries and mice, and they observe their behavior. For example, they observe if the fruit fly becomes more fearful or more aggressive after the gene manipulation. Machine learning comes into play because we generate thousands of hours of video that needs to be analyzed to identify if flies are being aggressive or fearful. We don't want a biology grad student to view thousands of hours of fruit fly videos. It is much faster and more efficient to use machine learning and related techniques to train a system that can automatically detect these types of activities from the video."

Yue adds: "Of course, there are commercial applications, as well. YouTube is trying to build a better search engine for videos. If you want to find a snippet of a certain action, you want to actually search inside the video rather than just tags of the video, which is what they do now. Video analysis has many other applications, as well, including tracking data for sports and tracking human motion to build realistic cartoon characters. This was part of what I worked on at Disney Research before coming to Caltech."

> *"We are trying to infer something about some quantity of interest that depends on an imperfectly known reality, and we turn this into an adversarial or Minimax game where the universe chooses reality and we come up with a model for it."*
>
> Houman Owhadi, *Professor of Applied and Computational Mathematics and Control and Dynamical Systems*

Houman Owhadi, Professor of Applied and Computational Mathematics and Control and Dynamical Systems, is also interested in estimating and predicting complex systems with limited information. "For instance, you want to predict the probability that the temperature of the planet will be in a given range in 50 years, but you have incomplete information about the underlying physical processes, limited computational capabilities, unknown probability distributions, limited data," he says. "How do you do that? Or there are systems that you have to build, but you are never completely sure about whether they are going to fail or not; nevertheless, you still have to make decisions about using them."

Making predictions and critical decisions with incomplete information is part of the human condition. An engineering label for the area is uncertainty quantification, which, as Owhadi explains, "is essentially a generic term which stands for a field that is emerging at the interface between probability, computational science and engineering, optimization, machine learning, and decision theory. There are many challenges in this field, and we can get an idea of what those challenges are by talking to people in the industry and in the National Labs. Oftentimes there is a need to answer very specific and critical questions, but the methods are not there. The mathematical methods have not been developed. Therefore, in this case, the application itself is driving fundamental research."

Owhadi continues: "We are trying to infer something about some quantity of interest that depends on an imperfectly known reality, and we turn this into an adversarial or Minimax game where the universe chooses reality and we come up with a model for it. What is most exciting to me is to use these techniques to guide, facilitate, or turn the process of scientific discovery into an algorithm. For instance, the question that I'm currently looking at is: Can the process of discovery of scalable numerical solvers be automated by reformulating the process of computing with partial information and limited resources, such as that of playing underlying adversarial information games?" He concludes: "These games can be difficult because the chessboard doesn't have 64 squares. It has an infinite number of squares, and calculus on a computer is necessarily discrete and finite. But, nevertheless, if we can develop a calculus to play on this chessboard, then we can turn the process of apprehending models of reality into an algorithm. So where can this take us? Everywhere!" ENG

Houman Owhadi

Visit cms.caltech.edu to learn more about the Computing + Mathematical Sciences department's faculty.