What do the complex workings of a cell have in common with the relentless unrest of the New York Stock Exchange? Are these dynamic storehouses of information obeying similar laws in terms of the ebb and flow of information, and can they be modeled, analyzed and understood by a unified mathematical framework? Caltech is launching a new Institute-wide intellectual adventure with the creation of the Information Science and Technology Institute (ISTI)—drawing the curtain back, so to speak, on the nature of information itself, and redefining the way we approach, understand, and implement information science and technology.

When You Come to a Multidimensional Fork in the Road, Take It:

Information Science and Technology at Caltech

Within the next decade, information at Caltech will be a unifying, core intellectual theme spanning the physical, biological, and social sciences, and engineering. Such a formidable, collective leap forward is the result of two idiosyncrasies: Caltech's longstanding and imaginative blending of traditional disciplines and the low one or two degrees of separation between disciplines, faculty, and students which allows exceptional people from seemingly disparate fields to work together naturally. Put another way: we're fabulously small, we engage in a lot of scientific gossip, and the standard departmental boundaries are all but invisible around here.

ISTI's interdisciplinary research, academic, and outreach agenda is large and will develop roots in each of Caltech's six divisions, with participation of more than 20% of the faculty, and nearly 35% of all students through curriculum. We aim to create a common language for the study of information, one that will stimulate fundamentally new thinking about problems facing not only the usual suspects (computer science, quantum physics, electrical engineering, applied physics, and applied mathematics) but also those not normally associated with information science and technology such as experimental economics, pure mathematics, and developmental biology. By approaching information science and technology from multiple levels of abstraction, we'd also like to figure out new tricks for atoms, light, molecules, cells, circuits, algorithms, and networks.

Jehoshua Bruck

What will be the outputs? Absolutely smashing scientific and engineering discoveries, students who'll go out into the world and (we hope) one-up their thesis advisors, and technological advances only yet imagined in our wildest dreams…

Such an ambitious program will involve two phases. The first is the creation of the research component, the wellspring from which the corresponding academic and outreach programs of the second phase will flow. And we need look no further than three words—multidisciplinary research center—for the formal mechanism for bringing people and their ideas together from each of the "six corners" of the Institute.

Over the last year, in an effort to define and help grow an IST community at Caltech, groups of faculty convened, conferred, and converged on a set of unifying principles for four new research centers that together provide the critical mass necessary to launch ISTI. Jehoshua (Shuki) Bruck, Gordon and Betty Moore Professor of Computation and Neural Systems and Electrical Engineering, chaired the IST Faculty Planning Committee, which issued its final recommendations in early January. The proposed centers, ultimately to be housed in a new building, are the Center for Biological Circuit Design (CBCD), the Center for the Physics of Information (CPI), the Social and Information Sciences Laboratory (SISL), and the Center for the Mathematics of Information (CMI). These four new centers will join the established Lee Center for Advanced Networking and the NSF Center for Neuromorphic Systems Engineering to form the initial core of ISTI. As ISTI matures, research advances and the natural dissolution of older research initiatives will drive the creation of new centers.

From these vibrant centers will emerge a unique academic program, the first of its kind in the country. The new undergraduate and graduate programs will combine engineering and science with a clear focus on information, and direct exposure to the central issues across the entire intellectual landscape. And finally, to create the broad societal impact commensurate with the outstanding research and academic components of ISTI, we will design and conduct a highly visible outreach program. Through executive, visitor, and industrial affiliate programs, we hope to supplement and share Caltech's contributions by collaborating with members from key academic institutions, government, and industry. Workshops, lectures, and summer schools will round out the menu for the continuing revolution in information science and technology.

Listen in on the following four conversations among Caltech faculty engaged in thinking about what these new centers will bring to Caltech and society at large, as Caltech embarks on this unparalleled and profound exploration.

Center for Biological Circuit Design: Soft Circuitry and Liquid Algorithms—A New Bioengineering Frontier Takes Form

A Conversation with Niles Pierce, Paul Sternberg, Erik Winfree, and Barbara Wold

Biology computes, that is, living structures store, process, and communicate information in organisms and ecosystems. The CBCD is being organized to understand the form and function of these biological circuits and to develop the tools needed to design new and improved circuits.

WOLD: There is a computing revolution going on across the board in many areas of biology—from molecular, to cellular, to developmental and neurobiology. At an obvious level, the revolution is driven by rapid changes in the kind and amount of data we work with, beginning with entire genome DNA sequences and everything that now flows from them. The basic challenge is to turn data into real information, then turn that information into real understanding. At another level, biologists have long been interested in information in living systems—how it is encoded, stored, recalled, and transduced from one site to another. These are themes that the faculty in this Center will be addressing in a very particular way.

After talking to many of our faculty and colleagues, in and out of the Biology Division, we hit upon the idea of focusing on biological circuit design. In some sense, you don't really understand the properties of something until you can sit down and—from scratch—design it, test it, and see if it behaves as you predicted: have you got it right? I'm not an engineer, but I think that's a major engineering process, or at least an important one. Biologists have been, so far, quite timid about wholesale design. We go in and tweak things a lot. We break things and see what happens. That's the heart of classical genetics. Or, we take things out of the cell and make them work in a test tube—that's biochemistry's challenge. So at this point, from all of our tweaking, the biologists have learned a lot about the molecular components of gene circuits. Similarly, neurobiologists know a lot about the cellular components of neural circuits that ultimately lead to brain function and behavior. In the middle are the people studying signal transduction—that is, how signals travel from the outside of the cell to the inside of the cell, or from one cell to another.

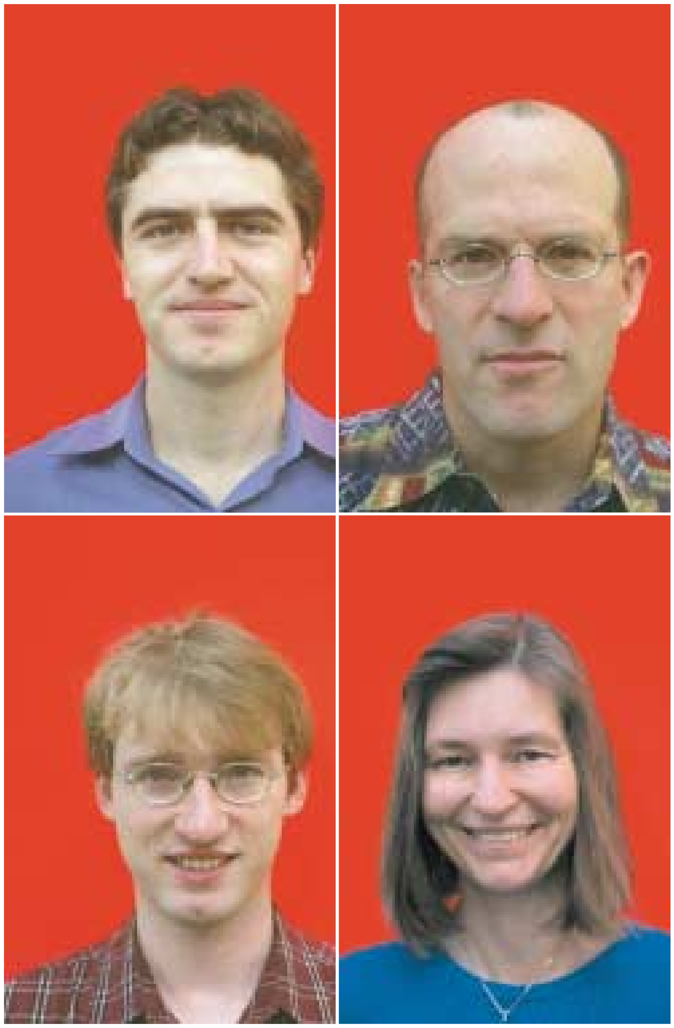
...you don't really understand the properties of something until you can sit down and—from scratch—design it, test it, and see if it behaves as you predicted: have you got it right?

The state of the art is this: we know an enormous amount about what the circuit components are, and something about how they're hooked together. We know a good deal about how the inputs work, and, globally, what the outputs are. But what gives a biological system its real properties—for instance, its robustness in the face of various kinds of insults? What are the dynamical properties of important circuits? How is information really encoded or stored by a given molecular or cellular circuit? Getting at these questions using a design focus is the core mission I see for the CBCD: it will take the fruits of all the research of past decades, combine it with the current revolution in biological information processing, and focus on circuit design. This is tremendously exciting, and central to deep understanding of biological systems.

STERNBERG: Another way to describe what we want to do is the "reverse engineering" of biological systems and circuits. But it's going to be much easier to learn how to do it with, for instance, a Model T rather than a Boeing 777. Organisms have been around for a billion years, making nature's designs incredibly complex and sophisticated. Even the simplest organisms are intricate integrated machines. They have embedded controls that are really hard to tease out. It's a lot easier to build something from scratch and then learn how to model it.

In my lab, we've looked at signal transduction and we've come up with very nice models that are powerful. But when we go into the real cell, they fall apart, because every little detail has been tuned by processes of evolutionary selection to make it work. That means you really want to start very simple. That's where the synthetic approach comes in. To build biological circuits, we need to define components and interactions. We have to determine which components are really going to be robust. You can liken this to creating a system using Lego bricks. You want to have the equivalent of those bricks, and that takes a lot of thought. Right now, Niles Pierce, Steve Mayo, Frances Arnold, and their colleagues are thinking about how to make those components.

PIERCE: We have three different types of people at Caltech who are all working in areas that contribute directly to progress on this very challenging topic. First, as Paul said, we have the tool builders who have been working on components. Steve Mayo's lab uses computational methods to design proteins with enhanced stability or novel functions. Frances Arnold's lab, by contrast, uses directed evolution to obtain molecules with new or enhanced functions. Richard Roberts' lab has developed a novel approach for *in-vitro* selection to screen for molecules with particular functions. My lab works on computational algorithms for designing molecular machines out of DNA and RNA. Erik Winfree's group is interested in biological computation and issues of how biological systems can be designed to process information. Steve Quake, Jared Leadbetter, and Frances Arnold are collaborating on the design of cellular signal-processing circuits in bacteria. Finally, we have a number of biologists [here and elsewhere] who study the structure and function of naturally occurring circuits, including Mel Simon, Elliot Meyerowitz, Stan Leibler, Paul Sternberg, Eric Davidson, Mary Kennedy, Thanos Siapas, Jim Collins, John McCaskill, Ron Weiss, Tom Knight, and Barbara Wold, among others. So there is a diverse set of people working on component-level issues for circuit design, creating synthetic circuits, or studying naturally occurring circuits. The latter have a deep understanding of how those circuits function and how they're structured. All three communities are well positioned, right now, to try to approach biological circuits from a synthetic point of view.



Clockwise from top left: Niles Pierce, Paul Sternberg, Barbara Wold, and Erik Winfree.

STERNBERG: There's a critical mass of talent, including researchers in the neural biology community—the Computation and Neural Systems program—who are thinking about how naturally occurring circuits work, and how one might like to design new ones. Because of the properties of the systems they study, they have a different view of how to analyze a complex circuit. Thanos Siapas and Gilles Laurent record information from multiple places in one structure simultaneously. They are good at articulating this approach and figuring out how to apply it to other complex systems—for instance, in a cell. Bringing in their expertise and interests allows us to make bridges all the way from chemical engineering to brain neuroscience in this quest to design and understand biological circuits.

Caltech is more than ready to make this very interdisciplinary, very ambitious goal happen. And again, it can only happen here, because in all the divisions you have people who are really good at what they are doing, of course, but also imaginative enough and interested enough to be able to learn other approaches. I think what will happen in the Center is that at the start, everybody will come in with his or her own ideas, leading to an incredible effervescence. Then we'll condense our focus on a couple of projects that seem tractable and seem to be the right way to learn to prove principles that will lead to new technology. The new technologies will then be applied in many directions and spawn new industries.

WOLD: One of the other things the CBCD will spawn will be an entirely new generation of students and post-docs with a worldview that is some interesting combination of all these inputs. Without the Center, a few students might make the interesting connections that biological circuit design requires. With the Center, and the concomitant "lowering of the energy barrier," so to speak, the path toward this kind of research training will be much more easily and frequently traversed. So the impact—through these people—ultimately goes far beyond Caltech.

WINFREE: One of the problems engineers face is understanding which aspects of a given component are important and which are just implementation details, not really relevant to the function. This leads to new levels of abstraction. For instance, we ask: "Is this atom over here the critical atom, so I need to focus my attention down here at the molecular level? Is it critical to the function or not?" Researchers try to understand that by making mutations, changing a moiety here or there, and so on. This is another approach for determining which parts of a system are important, and which are merely accidental. I hope that going through the design process will help elucidate this completely.

Caltech is more than ready to make this very interdisciplinary, very ambitious goal happen.

Many advances in biology have been driven by instrumentation. For instance, once people understood that a feedback loop coupled to an electrode in a cell could lead to something like a patch clamp, a whole new way to characterize biological circuits was born. It became possible to measure currents, I-V curves, and so forth. An entire range of experiments, previously impossible, became possible. The ability to build electronic circuits and integrate them with biology brought to the table a new way of doing science. The possibility of building *biochemical circuits*—for instance, novel genetic regulatory circuits to hold the concentration of an enzyme at a constant level, or to trigger a reaction just at the right time—will provide an entirely new approach for understanding what goes on inside the cell.

Programming with biochemical reactions rather than with logical "and gates" and "or gates" is a whole different beast.

One thing that excites me is thinking about the relationship between the concepts that computer scientists have developed and the realities that biologists are observing. Understanding the kinds of algorithms that biology has been exploiting, and the design space of those algorithms, is fascinating. Programming with biochemical reactions rather than with logical "and gates" and "or gates" is a whole different beast.

ENGENIOUS: What are the practical goals of the CBCD?

PIERCE: One way to encapsulate a long-range objective for the CBCD is to say that we're going to try to recreate the remarkable technology of the compiler. A compiler takes an algorithm written in a programming language and turns it into instructions that a computer can understand. Given a conceptual design for a circuit, we'd like to be able to "compile" a set of molecules that can be introduced into a test tube and be observed to function according to the principles for which that circuit was designed. This outcome would be tremendously exciting not only for its biotechnological and medical applications, but also for the sheer challenge of working with a complex array of components to develop a design framework robust enough to produce working molecules and circuits. This goal sets a high standard, but I think we have a real shot at meeting it.

ENGENIOUS: Will principles of evolutionary biology be useful in this work?

WINFREE: Exploiting evolutionary principles in the design process is already being done at Caltech. For instance, Richard Roberts does *in-vitro* selection to design protein sequences with functional properties. Frances Arnold applies directed evolution to both circuits and proteins. These are important tools. It will be interesting to integrate this "irrational" approach, where you try a bunch of things and select one that works, with rational, systematic design, where you put together a system based on your ability to predict how it will function.

WOLD: A hybrid approach is to design first, then subject the system to very rapid evolution for optimization. This allows you to see how close you were to optimal in the first place.

WINFREE: Absolutely. I think that's an important approach, and the way you might design components—a particular protein, for example—by some kind of directed evolution, then characterize it, put it in your toolbox, and fit it into a circuit in a rational way. Then, perhaps, do another level of evolution to optimize that circuit.

STERNBERG: Then you can look at evolution to see what's worked—which components have been used

> *...we're going to try to recreate the remarkable technology of the compiler.*

in many circumstances, but have maintained their central character. Neurons, for instance, are very successful. Our neurons are the same as many other creatures' neurons, but they're wired together in different ways. It's the circuit design that makes us different. That's something that was discovered at Caltech, by John Allman and his colleagues, and elsewhere. Neurons are one type of component. At the molecular level, we have the G protein, a molecule that Mel Simon has been obsessed with for years. It

> *What Caltech does best is getting fundamental ideas and technologies to a level where they can radiate out to the tech sector.*

acts as a little molecular switch or timing device. We could start with these known robust components and learn how to build things with them. But given the collaborations that will take place, each person's research approach might be wonderfully and radically changed.

WOLD: That may be the most important bit of "evolution" from our immediate point of view. We exert intellectual pressure on each other to look at a problem in a different way and to use somebody else's point of view—intellectual evolution in action.

STERNBERG: And that's why articulating and committing to a focus on designing circuits is going to change the direction of many people. There are a lot of our colleagues we think will be involved, but they don't even know it yet…

WOLD: But we have faith that they will be attracted by the theme and know exactly what to do.

One last thing concerning potential practical outputs. Our greatest passion is for the deep underlying principles. At the end of the day, to us, a practical result would be having the compiler. But as viewed by many other people, that's not a practical output. Certainly the implications of this work will have a significant biotechnological spillover. What Caltech does best is getting fundamental ideas and technologies to a level where they can radiate out to the tech sector.

WINFREE: And the possible technological implications here are not restricted to the medical or biological realm. The ability to program things and to automate tasks has profoundly affected science, engineering, and technology in the last 50 years, a very short time historically. Most programs exist in microprocessors, which are quickly becoming ubiquitous in our lives. They are in your microwave oven, in your car, in your digital camera. We know how to program and exert embedded control over macroscopic electromechanical systems, and this has revolutionized technology.

Nature, through biological processes, has transformed the earth by exploiting algorithms and embedded control at the chemical level to fabricate cells, bodies, and ecosystems; to build forests from light and chemical nutrients, for example. Intellectually, we don't really understand how these things exert an influence over chemistry and organize it into meaningful constructs. Biochemistry is where we see most clearly that information and algorithms are fundamental elements of the chemical process. Nature has polymers, like DNA, which contain information. The cell interprets that information as a program for directing its behavior. Evolution changes the program to carry out an incredibly wide range of functions. This is a technology that isn't just biological: biology is only one possible result of programming biochemistry. Working with atoms and molecules in systems will turn out to encompass a wide world, and is going to be very fun.

WOLD: Actually, at the end of the day, that's the point. We don't usually start with that. What's the goal of your Center? To have fun. But we know it will be… E N G

Niles A. Pierce is Assistant Professor of Applied & Computational Mathematics. Paul W. Sternberg is Professor of Biology and Investigator, Howard Hughes Medical Institute. Erik Winfree (PhD '98) is Assistant Professor of Computer Science and Computation and Neural Systems. Barbara J. Wold (PhD '78) is the Bren Professor of Molecular Biology and Director of Beckman Institute.

Center for the Physics of Information: The Impending Overthrow of the Silicon Monopoly: Revolutionary Substrates Unite!

A Conversation with André DeHon, John Preskill, and David Rutledge

Silicon is a superb computational substrate...but sooner or later it will run out of room. The CPI is devoted to inventing the new computational substrates, architectures, and algorithms for the computing devices of the future.

From left to right: André DeHon, John Preskill, and David Rutledge.

PRESKILL: We're kind of an odd mix of people, you know. I'm a theoretical physicist, Dave's an electrical engineer, André is a computer scientist. But I think we have some things in common. In those areas of overlap there's a potential for some really exciting scientific and technological developments. We know that the advance of our information technology, which has been dazzling for so long, is confronting limitations that come from physics and, in particular, from the size of atoms. And we don't know beyond say, a decade, what we are going to do to continue the type of progress we've gotten accustomed to. It's going to require really new ideas. We don't know what. We don't know how we're going to get there. And that's what we're going to be thinking about in this Center. There are a lot of ideas about exotic ways of manipulating information, but there's a tremendous gulf between some of those concepts and practice. In particular, I'm interested in quantum computing. If it comes to fruition, we'll see an amazing advance in the speed of computation. It's really exciting. We have these beautiful theoretical ideas about quantum computing, but we really don't have any definite idea about how to progress along the road that will lead us to advanced quantum computing.

RUTLEDGE: One thing that I think is interesting about the Center is its ancestry, so to speak. Caltech has a very good history of making fundamental contributions to the physics of small things and information. Three people that come to mind are Richard Feynman, John Hopfield, and Carver Mead. There's a great tradition. But recently Caltech has hired many outstanding junior faculty in different departments across the campus who are connected to this area. That's really Caltech's advantage.

We have the opportunity here to take some of the ideas being developed on the scientific, physics side to see if they really work in engineering products. That would require, for example, getting some of the ideas to work on a silicon integrated circuit. This vertical integration—from the theoretical up through the practical—will mean strong collaboration between scientists and engineers to get really neat scientific ideas transformed into practical devices.

DEHON: I think vertical integration on a higher level also means we'll be rethinking abstractions at many layers. Presently, we've got a very well-developed set of abstractions for designing computers and software on top of silicon. And we know "this is where we collapse into the gate level; this is where we build up some architectures on top of that; here is where we build the program; and then there are algorithms on top of that." There is a nice set of defined layers. On the other hand, when the rules change, the costs change, and really good engineers will be the ones saying, "Okay, these old abstractions are getting in my way." What's very clear here is that using some of the same interfaces and abstractions we have in the past will defeat the purpose. Silicon's been very reliable; things work because we're talking about a million atoms sitting in one place. But it's not clear whether we'll have that type of control with substrates where we will be working with individual (or very few) atoms. So that's going to force us to re-evaluate all of our models: what you use for computation, the programming language, and so on.

...the advance of our information technology, which has been dazzling for so long, is confronting limitations that come from physics...

...really good engineers will be the ones saying, "Okay, these old abstractions are getting in my way."

ENGENIOUS: So for instance, algorithms might not sit so high in the hierarchy any more? They might be more embedded in the fundamental substrate itself in some way?

DEHON: I would believe that.

RUTLEDGE: Also, with smaller numbers of atoms, you really have to deal with errors in a fundamental way…

DEHON: …because there are some things that may be less hidden. One of the things you try to do in good engineering abstraction is hide unnecessary detail and bring up the dominant effects you need to optimize. I think the dominant effects are probably going to shift and change. There are different things we'll need to bring to the attention of the engineer.

PRESKILL: Maybe the concept of a general-purpose device will be less central than it was in the past. Some physical systems may be better suited for certain applications than others. We should be willing to let blur those layers which had served us very well in the past—substrate, architecture, and algorithm—and to think things through from the start. Error correction is probably the best example. In quantum computing, this area is one of my major interests. For instance, we had to rethink what type of physical system would potentially be very resistant to errors. Some technologies with lots of good features may fail in that regard. So quantum computing just won't be a possibility for certain types of physical applications.

DEHON: The deeper I got into the VLSI work I started out with, the more I began to really understand that the underlying physics of the substrate was inseparable from the most efficient architecture possible, and the eventual implementation.

And as VLSI got smaller, the landscape changed. Wires got more expensive, for instance. Ultimately, our computations do depend on the physics we use and the structure of the physical world. After looking down at VLSI for so long, it's good to just look up and realize scientists are working with some amazing new phenomena: carbon nanotubes, experiments trapping a single atom. So you say to yourself, "How can we harness these things?"

For me, a central issue is understanding computational cost structure. When the cost structure changes things radically, the nature of the solutions changes as well. The general-purpose processor that made a lot of sense in VLSI just doesn't make sense for these new things. We are off in a completely new playground, which is very exciting for an architect. Caltech is a place that allows me to think sometimes at the circuit level, sometimes at the manufacturing level, sometimes at a mathematical/statistical yield level—all over the map. And for something new like this, where no established discipline exists, it's important to gather people from various areas who can think broadly about the issues. This is what the CPI will accomplish.

PRESKILL: We're searching for new paradigms, something that Caltech does especially well. Maybe we won't be the place that actually builds the next revolutionary generation of devices, but I think what we should aspire to is becoming the world's leading institution for laying the scientific foundations which will be the basis for information technologies of the future—we will be generating absolutely new ideas. And training students so they have the broad background that's necessary to get the big picture.

ENGENIOUS: How will the structure of the Center facilitate breakthroughs?

RUTLEDGE: We're interested in creating an environment conducive to professor and student interaction. And we're anticipating that there will be a new Information Science and Technology building as a result of the fundraising campaign. University professors are prone to being trapped in an area; this is a good way to force them out into new things.

...with smaller numbers of atoms, you really have to deal with errors in a fundamental way…

DEHON: People like Bill Dally [PhD '86; now Professor of Electrical Engineering and Computer Science at Stanford University] and others came to Caltech in the early '80s because it was *the* place for VLSI. And that's really what we want—for Caltech to be *the* place for the next revolution in novel computing. There's a great deal of uncertainty about what's going to happen in this area, and yet that's what makes it exciting. What's going to happen at the chemical level? At the biomolecular level? At the quantum level?

It's so important to have the freedom to be daring...

Look at this from a student's perspective. I maintain that our current and future students will go out into the world and have the same impact the Caltech VLSI students are having now—maybe even more so if we can get students from every area to interact with each other. For example, a student comes here to study molecular electronics, but this area doesn't exactly pan out. However, the real benefit will have come from interacting closely with other people doing perhaps biomolecular and quantum work, and from being taught how to think broadly about these areas. I think our students will certainly be in a position to found, transform, and lead the industry.

PRESKILL: The students are really the key. Caltech should be *the* place, the number one place, that a student thinks of if he or she is interested in the future of information technology in the long-term. Actually André and Erik [Winfree] did a great thing this summer—they were involved in the Computing Beyond Silicon Summer School, which attracted people from all over.

DEHON: We had 45 students for four weeks and 12 guest lecturers—the top people—coming from different institutions and intellectual areas. It was really something.

ENGENIOUS: How did the students deal with this new conceptual framework?

DEHON: It was interesting because it's not a "done thing," there is no orthodoxy. The students

[continued on page 27]

Computing Beyond Silicon Summer School

Or How I Spent My Summer in Pasadena

There are revolutions at hand in the way we understand an implement computation, driven by an awareness of impending barriers to VLSI scaling and new understandings of the physical world. This fundamental shift in perspective allows us to contemplate engineering computational substrates at the molecular and atomic levels. To develop and exploit these new substrates will require an intimate understanding of both the physical substrates and the nature of computation, as well as the relation between them. Research and researchers whose competencies span across the disciplines will be necessary to drive progress in this area of novel computational substrates…

…Thus read the opening paragraph of the announcement for the Computing Beyond Silicon Summer School (CBSSS). Coordinated by André DeHon, assistant professor of computer science, and Erik Winfree (PhD '98), assistant professor of computer science and computation and neural systems, the program brought together leading research faculty and 45 outstanding undergraduate and graduate students from many disciplines and institutions across the country (including Caltech). Part boot camp, part pleasure cruise, CBSSS served as an intensive four-week introduction to the emerging fields of molecular, biomolecular, and quantum computing.

Lectures, reading assignments, and a paper and presentation project kept the students active. In between all this, students seized the opportunity to hang out with the guest lecturers, Caltech faculty, and each other. They came, they learned, they met future collaborators—and they had fun. A potent combination. And of course, ditto for the faculty and guest lecturers…

As a prototype of ISTI's outreach program of summer schools, CBSSS's unique collection of people and ideas in one place at one time points to the future of Caltech as a hotbed of research in novel computational substrates.

For more information on who was there and what they did, go to
http://www.cs.caltech.edu/cbsss

The CBS^3 students gracefully posed for "mug shots" for posterity. To engage the students beyond the lectures, the CBS^3 faculty asked them to self-organize into small project teams to expand on issues related to or motivated by the subject matter presented in lectures. The students had roughly three weeks to focus in on a topic and put together a brief report. See http://www.cs.caltech.edu/cbsss/report1.html for the resulting collection of student reports. Almost none of the students were "experts" in the issues they studied when they entered the program. Nonetheless, these reports show that the multidisciplinary teams assembled were able to dig deeply into a number of interesting problems and point out some promising directions for further inquiry.

[continued from page 25]

definitely went through a little mind expansion. There were EE students who thought [the EE framework] was the only way the world works...and in some cases biology students who didn't at the outset realize that maybe computational complexity meant something to them. All of them were challenged and out of their comfort zones. I think many of them had the experience of "Wow, the world is bigger than I thought it was." There is an opportunity to do interesting work at, for instance, the intersection of computer science and biology.

PRESKILL: And in some ways, it's easier for students than it is for us, you know. For me, the work I do at the interface of physics and information science seems kind of "out there," novel and daring. But to my students, it seems very natural. Those are the things they're interested in. Combining computer science and physics is second nature for them.

ENGENIOUS: Caltech seems to have both a deep intellectual reservoir and a smallness of size that allows us to attack these problems much differently than anybody else. Are there other universities that can do what you anticipate doing?

RUTLEDGE: Smallness is a part of it. Caltech feels the same size as the entire EE department at Berkeley. There, someone "far away" from you intellectually meant someone that was making superconducting detectors in the electronics department. However, there are a lot of good places out there, and a lot of competition.

DEHON: Certainly MIT has the breadth. On the other hand, it's a big place—with something happening in the Media Lab, and then there are people over in the AI Lab, far from folks in Mechanical Engineering. So you know, maybe it's a little bit harder to get coherence between the groups.

We are off in a completely new playground...

ENGENIOUS: What is the one thing that excites you most about the Center?

PRESKILL: Well, from my own parochial point of view, I'm excited about making quantum computers a reality. It's just one of the emerging frontiers. If something like the Center for the Physics of Information can make that possible, I think that's very exciting.

DEHON: The Center will really allow us the opportunity to build critical intellectual mass. My students and I can sit there and ask each other questions, but having the ability to work with people from other areas thinking about the same problems will be powerful. The new solutions will create new abstraction hierarchies and new ways of decomposing problems. Things will not be the same as they were. Let's think out of that proverbial box and come up with some wild ideas.

RUTLEDGE: I see two things. One is the opportunity to work with people across a wide range of disciplines in a serious way. And the second is consistent support. I've run government centers, and it's astonishing how much of your life gets taken up by requirements and crazy things that change right in the middle of established projects. Just to get out of the kind of environment where you're told what to do every two months is liberating.

PRESKILL: Absolutely. It's so important to have the freedom to be daring, not to have to defend the project on the basis of some short-term goal, some milestone event.

RUTLEDGE: I want to mention that junior faculty will be instrumental; they have already contributed in fundamental ways to getting things started. People like Erik Winfree [PhD '98], Ali Hajimiri, Hideo Mabuchi [PhD '98], André of course, and a handful of others.

PRESKILL: Yes, I think that's pretty good evidence that we're on the right track. Looking around campus and seeing so many young faculty involved in exciting projects at the interface of physical science and information science tells me that we are in a good position to live up to the legacy of Feynman, Mead, and Hopfield. E N G

André DeHon is Assistant Professor of Computer Science. John P. Preskill is the John D. MacArthur Professor of Theoretical Physics. David B. Rutledge is the Kiyo and Eiko Tomiyasu Professor of Electrical Engineering.

Social and Information Sciences Laboratory: Markets and Other Noisy Human Artifacts—Can Computation Bring Them Out of the Bronze Age?

A Conversation with Yaser S. Abu-Mostafa, K. Mani Chandy, and John O. Ledyard

Social systems such as financial markets, political processes, and organizations aggregate and disseminate immense amounts of noisy information—but can this be done more efficiently? And can new, innovative structures be invented with the assistance of more sophisticated information technology? SISL will be exploring these and related issues.

ABU-MOSTAFA: There is an abundance of data and an abundance of computational resources in the world, yet our ability to manage these resources, to be able to look at data and efficiently extract the correct information, is limited. Highly distributed, data-rich, and generally unstructured, the world's financial markets seem to work well—remarkably well given the loose structures and lack of supervision—but they can be improved. The players in the markets are individuals, institutions, sometimes simply computer programs. They are looking at pieces of information that may be different from one source to another. They're all interpreting information differently. They have their own ideas and preferences regarding risk, value, volume, etc. Eventually, all of this is aggregated in global quantities like price, volatility—things of that sort. So a basic understanding of how such a general system results in efficient information aggregation is very important for two fields: economics and engineering. On the economics side, we would like to better understand markets and eventually be able to design markets. Once we do that, we can design markets in different arenas where there are no markets now. From the engineering perspective, we're interested in learning from the principles of how markets work how to generally manage distributed information and be able to aggregate it in a meaningful way.

LEDYARD: Economists would suggest that perhaps they know something about markets already, that 200 years of study have produced remarkable insights about them. What's of importance in this Center, however, is the role of technology in the way markets operate.

The question is whether we can leverage new advances in information science and technology to design new markets.

There are barter markets, which have been around for thousands of years, which are not very efficient. The information technology underlying the New York Stock Exchange is still primitive in that humans are crucial at many points in the process. Many aspects of markets work wonderfully. If I'm fixing my house and I need a nail, I know I can go to the hardware store, and the nail is sitting there waiting for me. How did they know I would need a nail that day? It's not centrally planned. It's not managed the way engineers like to manage things. It's dispersed, disorganized, decentralized, but it does compute some pretty incredible things.

There are other pieces that don't work very well: supply chains, for instance, and public good kinds of problems. Markets don't work very well in these cases, partly because there aren't very many participants. They're very specialized and may not have much volume, so you can't rely on immediacy. The question is whether we can leverage new advances in information science and technology to design new markets. Economists have generally attacked these problems assuming computation was free and easy—which it's not. Bringing the reality of information processing into market design is really important. The role of SISL is to bring the expertise of engineers and information scientists together with the expertise of economists—each has something the other doesn't. Working together, something really special will emerge.

ENGENIOUS: Will you be inventing new computational tools to deal with these problems?

CHANDY: At this point, I don't think we really know. That's why SISL is so interesting. From my point of view, the research of this center will bring "power to the people." Economic power has two parts: resources and information. Information technology today is at a place where one half of the economic power equation—information—is widely available. And this represents a significant dispersal of power from the few to the masses. I'll give you three examples of how this is going to change your life.

When the defense department wants to buy planes, it puts out a request for proposals, companies respond, and they finally choose a plane. DOD can afford to do that because DOD budgets billions of dollars for a plane. If you want to buy a car, you don't have the same flexibility. You don't request proposals for cars that fit your specifications. Nor, if you want to travel, do you put out a request for proposals for tours with certain specifications. You can't do that because the cost of the transaction is high. But apply computational resources to this scenario, and things will change dramatically.

The second example is futures markets. We are familiar with the futures market on things like wheat, oranges, pork bellies, and so on. But what if there were a futures market on services like carpentry, plumbing, and electrical work when you add on to your house?

The third example is the creation of financial derivatives. Today, large financial services companies create financial derivatives tailor-made for companies doing ship-building in Poland, for example. Financial services companies create custom-made derivatives and sell them for lots of money. But with the kind of technology we will develop, companies will want to sell you derivative products for yourself based on your personal situation.

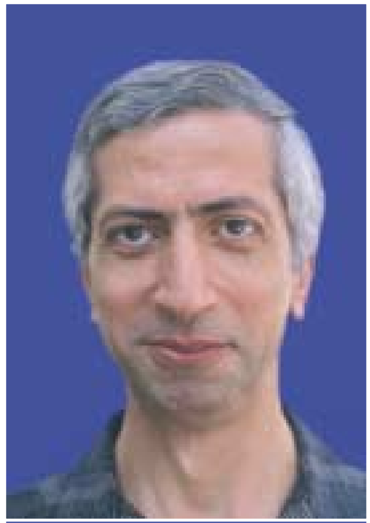
So these are a few examples of how the Center's research will help economic power devolve to "the people."

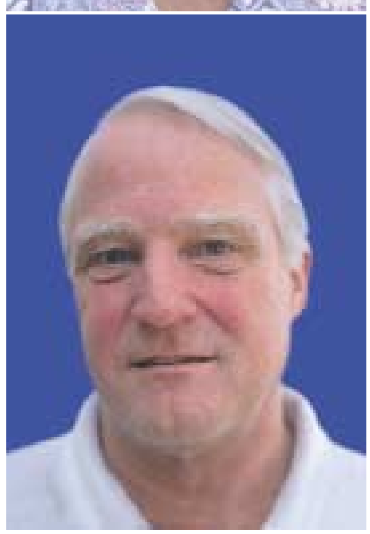
LEDYARD: Here's a sort of common theme in the story: let's say you want to build or buy something, a car or house or vacation. Today, you have to go to somebody who's packaged everything up without your particular needs or desires in mind. You can have people specially build your cars for you, specially build your house, but it's expensive. With computational capability, you can allow people to express what they really want to buy in a marketplace. So, rather than hiring a project manager to build your house, the computer organizes schedules, locks in the futures contracts on carpenters, masons, roofers, and locks in a schedule. This is going to require some interesting theoretical work in terms of how you capture what are essentially "metaphorical" ideas—the idea that I want a house overlooking a lake, with three stories, etc.

The classic example of where this gets mishandled is the California electricity market. That was a designed new market. Somebody said "Let there be markets," and *voilà*! They did that in Russia and it was a disaster because they forgot they needed banks and property rights and various other things. In California, they forgot to integrate engineering, electricity, and the laws of physics with the market. They also made some bad assumptions about how people behave. There's been research, a lot of it at Caltech over the last 30 years, which could have prevented this problem from occurring. Simon Wilkie had a very nice article in *Engineering and Science* [Economic Policy in the Information Age, *E&S*, Vol. LXIV, No. 1, 2001, page 28] on just this problem. Engineers like to control everything. Economists hate to control anything. Integrating these two kinds of approaches is going to be interesting, but it's required for a successful energy market. Give SISL up to ten years, and we'll pull it off.

Designers of distributed systems can control the rules of the game, but they cannot control the players.

With experimental economics, we have a way of demonstrating to





From top to bottom: Yaser Abu-Mostafa, Mani Chandy, and John Ledyard.

people how these things really work. We can actually bring science to bear on it. The combined energies of those working in this Center will create an intellectual core that anybody working in these fields simply won't be able to ignore.

ABU-MOSTAFA: I'd like to add something to the idea of exotic derivatives: one of the biggest advantages of having the computational technology to price these things is being able to communicate the derivative to so many players, thus creating a commodity. It becomes a real market—a place of exchange between buyers and sellers—because of the number of players and because of their ability to come to an agreement on price and to communicate instantly.

CHANDY: John said that engineers like to control things… but a true distributed system is one in which you don't know the participants, or even how many there are. Designers of distributed systems can control the rules of the game, but they can't control the players. So there are two parts to a distributed system: the visible hand, or the rules by which all the participants play, and then the invisible hand—how many participants, and how participants operate provided they play by the rules. Markets are beautiful examples of this, and we need to understand better how we get global behavior from these policies. This is very much an engineering problem.

Engineers like to control everything. Economists hate to control anything.

LEDYARD: The process Mani is describing is what economists call mechanism design. It's also very much an economics problem, where we recognize the incentives people have to follow the rules or not.

ENGENIOUS: What other Caltech faculty do you anticipate being involved?

CHANDY: In computer science, there are two relevant areas: applying economic principles to distributed systems, and applying technology to economic principles. For the first part, we have Steven Low's work on the internet and algorithms, and also John Doyle's theories on control and robustness applied to non-traditional applications like markets.

LEDYARD: We have been using markets as examples, because many people have contact with markets. But the same conceptual structures and questions arise in issues of voting and elections, committees, and organizing large organizations. In my Division, we have Tom Palfrey working on political processes. Peter Bossaerts studies the dynamics of financial markets and the process of price discovery. Charles Plott studies information aggregation processes. Matthew Jackson does fundamental research on networks. All of them will be involved, as well as others.

CHANDY: We will also work with people from the Center for the Mathematics of Information. We share an interest in the growth of data, the extent of data. Essentially, data come in three forms. There are structured data, like the price of a car. There are totally unstructured data, like news about an explosion in Azerbaijan near an oil well. And then there are semi-structured data, for instance, auction information like you would find on E-Bay. All three kinds are increasing everyday, so the work of that Center—creating efficient representation choices—will be useful to the work of SISL.

ENGENIOUS: This work, taken as a whole, sounds like it could be an entirely new intellectual discipline.

If we knew what would happen two years from now, it wouldn't be research.

LEDYARD: It has the potential.

ABU-MOSTAFA: When you design a research enterprise like this, you have to have a gut feeling about it being special. But then these things create a life of their own. If we knew what would happen two years from now, it wouldn't be research. Once the collaborations begin, who knows what can happen? We've been discussing markets because they are tangible, and have real and immediate impact on people, but there is a wide range of applications for this research, including the organization of corporations, the health-care system, etc.

CHANDY: I really believe that SISL will have a direct impact on society, on ordinary people in addition to large institutions. This confluence of economics and information technology will impact everybody. ENG

Yaser S. Abu-Mostafa (PhD '83) is Professor of Electrical Engineering and Computer Science. K. Mani Chandy is Simon Ramo Professor and Professor of Computer Science. John O. Ledyard is Allen and Lenabelle Davis Professor of Economics and Social Sciences.

Center for the Mathematics of Information: Information Theory Revisited: Mathematicians and Friends Tackle the Whole Enchilada

A Conversation with Emmanuel Candes, Michelle Effros, and Pietro Perona

Mathematics has provided the foundation for virtually every major technological advance of human society. And now, there is a fundamental need to rethink the meaning and scope of computation, information gathering, and extraction. CMI will provide a home to the dedicated community of mathematicians, engineers, and scientists concentrating on developing the key mathematical ideas necessary to take information science forward.

From left to right: Emmanuel Candes, Michelle Effros, and Pietro Perona.

EFFROS: There is a lot of excitement in research at the boundaries between traditional areas. The thrusts of ISTI reflect that excitement. One way to cross traditional boundaries is to focus on the applications of information science. Another way is to work on the topics that different information science applications share—which will be our approach. The CMI is focused on understanding the essential nature of information itself, the common properties shared by information in all of its physical forms and applications. We hope to learn how to collect, quantify, communicate, and manipulate information efficiently. In studying the mathematics of information, we will bring together mathematical tools from communications, statistics, signal processing, and computer science with those developed across a wide variety of applications and build a shared foundation for studying information science.

Over the past 50 years, practical problems in communications, controls, and electronics have benefited enormously from breakthroughs in mathematics. The job in the information sciences is by no means done. Roughly, communications looks at bandwidth, controls looks at feedback, and computer science looks at computation. What is needed for today's more complex systems, whether natural or designed by people, is some way of capturing these things together and understanding how they interact.

Representation choice is one example of an area to investigate. Imagine that you have raw bits of data, or raw signal, and you want to extract from that some core meaning. Many fields have looked at the question of how to go from raw signal to information, but so far none have entirely automated the process. Humans are still critical in extracting meaning from data. Whether it's patient statistics collected by the Center for Disease Control in an attempt to identify epidemics early, or weather patterns tracked by the National Weather Service to warn people about impending storms, or genetic information gathered by researchers trying to understand patterns associated with heredity and disease, the quantities of information are enormous and the need for people to be a central part of the information extraction process is a critical bottleneck for advancement.

PERONA: The more we are able to dig into data and make sense of it, that is, transform data into information, the more powerful we become. The more efficient these processes are, the better we can make all kinds of important decisions—medical, economic, technological, and so on. Humans are built in a way that they spontaneously try to organize information and make sense of it. But machines are not built this way. There is an amazing amount of clutter out there in the world. We need to find out how to automate this process of easily understanding which features are the important ones—and which to ignore.

CANDES: Humans use representations all the time. Look at the history of simply expressing numbers. The Romans came up with a numeration system, but they had to give it up because it was not really efficient for calculating. If you try to add two numbers in the Roman system, it's a complete mess. That's why the Arabic numeration system was adopted, because it's handier to perform more complicated tasks. Now we have digital computers that use a binary system—only 0s and 1s—which makes addition, subtraction, multiplication, and division easy. This concept of representation is really critical to scientific thinking. For a given problem, you really want to find the correct representation—the one that makes a set of specific tasks completely trivial.

There is...a fundamental need of rethinking the meaning and scope of computation, information gathering, and extraction.

PERONA: Representations are not self-contained, they are finalized toward certain tasks. On one side we have the data, on another side we have prior knowledge about the world, and on the third side of the triangle is the task. All three determine which representation should be used for a given problem. This is one of the big themes for the Center. For instance, my colleagues here at Caltech are studying the brain's different representations of the physical space around a person. Photons create an image that is captured by the retina, and then objects in the image are assigned retinal coordinates. Next the objects are expressed in head coordinates, and then in body coordinates. All of these different representations are useful. If I move my eyes, I want to know where the object is in respect to my head or my body, because my eyes have to move with respect to the head but I want my representation to be invariant with respect to that motion. If I move my hand to rub an object, the object has to be represented in world coordinates so that I can find it both with my hands and my eyes. The brain makes at least two different versions of geometric representations of the world. We don't know for sure that these representations are cartesian either. The problem is made more complex in that there may be several representations of the same data that need to be coordinated—this is another big theme for the Center.

Attention and awareness is another related problem—organisms pay attention to only fragments of the sensations transmitted to the brain, because it is the most efficient way to operate. When confronted with practically infinite data, how do we know what to pay attention to? How do we shift our awareness? Several researchers in the Computational and Neural Systems option are dealing with the engineering issues behind awareness and will play a big role in the CMI.

EFFROS: Many people on campus are focused on representation choice. Some are concerned with vision, some with attention and awareness questions, and we have computer science people thinking about representation choice for the purpose of being able to do certain kinds of computations. The CMI will bring all these electrical engineers, computer scientists, and applied mathematicians together to tackle the foundations, the fundamentals of representation choice independent of the realm of application.

CANDES: I'd like to emphasize the timeliness of the Center. It's clear that scientists and engineers are engaged in acquiring massive data sets—in many areas of biology, bioengineering, and finance, many people are involved in massive data collection. It's clear that any kind of progress we make in the area of data representation will have a huge impact across many sciences. And though we're not the first ones to think about data representation, we do feel that existing representations are somehow limited. There's a whole world out there of new representations that we would like to explore systematically. Any major advances that we make will be useful to other key players in the other ISTI centers.

EFFROS: What is the smallest amount of computation I can use to perform a particular task? My own field of communications or information theory focuses primarily on the quantity of information, whether you measure that as bandwidth or just as the number of bits that you need to represent some particular piece of information. Controls researchers focus on feedback. To think about how these different resources interact or trade off is fascinating to me. If I'm working on a control system, say a distributed control system where I have a bunch of different devices all trying to work together to perform a particular task, I care about many things. I care about how many computations each one of them needs to do separately. I care about how much communication between them is necessary to make the system work. I care about how best to use feedback. I care about representation choice. And I care about all of these things *simultaneously*. We are now at a point where it is, I believe, critical to figure out how to put all of these pieces together. So in information theory, the traditional view has been to look at how many bits it takes to communicate or store information, but the computation resource has been considered to be unlimited. You can have as much computation and delay as you want, but feedback is going to be a problem. These other resources were allowed to be unlimited so we could see where the critical points were in the one resource on which we focused. If you look at these other fields, they've done the same kinds of things. However, researchers in each of these fields are now realizing that we really need to take all of the resources into account.

There are people all over campus who are thinking deeply about the mathematics of information. The goal in many senses is to bring them all together.

Taking advantage of Caltech's small size and cross-disciplinary nature, we think that we can make real progress in putting these things together. In trying to understand, for example: is there a dynamics of information? What would the dynamics of information look like? Is there a conservation of information? What are the properties of this new resource of information? Do they parallel the properties that we have in the physical sciences?

We have been building on Shannon's work for 50 years. But Shannon made assumptions. He did not constrain the amount of computation and he did not constrain the amount of delay. He just said, "Let's look at how many bits it takes," either to communicate through a noisy channel, or to store information. He captured one resource with incredible clarity and beauty by abstracting away many of the other resources that are critical.

ENGENIOUS: Could this Center reinvent the fundamentals of information theory, in a sense?

CANDES: If you allow me, I would like to formulate a more modest mission.

Every single field of scientific research is called upon to develop novel tools to process the information contained in massive datasets. While many aspects of these advances are going to be field-specific, it is clear that these challenges cannot be answered only in a peripheral manner. There is, in fact, a fundamental need of rethinking the meaning and scope of computation, information gathering, and extraction. From many endpoints of scientific research comes the solicitation to redefine our approach to information processing. Such fundamental paradigm change can, however, happen only if we invest a considerable amount of resources in theoretical thinking centered around information. In short, our Center will create an environment, a home if you will, where these things can happen.

...this is the place where we destroy all the boundaries between disciplines and even the concept that the disciplines need to exist...

First, the Center will create the opportunity to deploy mathematical ideas, theories, and algorithms in information technology; to import new challenges into mathematics; and to create new mathematical theories and new mathematical tools via these interactions. Second, the Center will strengthen existing interactions and create new bridges between mathematical science and key areas in information technology. And third, the Center will help train a new generation of scientists in this emerging interdisciplinary area.

EFFROS: It's not that there's something wrong with the pieces that are there. But it's as if we have a few pieces of the puzzle that only give us focused pictures in certain realms. We're missing the big picture that puts it all together into a unified whole.

PERONA: You could take a more top-down view and notice how, in the past century, technology delivered systems that were extremely effective at doing one thing. Think telephones, personal computers, automobiles, airplanes. All of these things are well designed and deliver the goods. They have changed our lives. Nowadays, things are being integrated and connected so you have telephone sets that become PDAs and computers; and automobiles that include telecommunications. And this is just the beginning of ubiquitous networking. These systems are increasingly complex. However, they're completely stocked with software that was designed 30 years ago. Unfortunately, we don't know how to design these integrated systems; we cannot guarantee that they will be robust to viruses and software glitches or that they will be stable and will perform according to plan.

A big theme in this Center is coming up with key mathematical ideas that will allow us to think about large, complex, distributed systems that include computation, include control, include communications, and still be able to deal gracefully with the inevitable software bugs, hardware problems of all sorts, and human errors. They have to keep working. Humanity depends on these systems. We are far past the point of simply needing the water well and the chicken and a tree hanging with fruit to live. If the internet goes down for a week, I think the world will stop. So the design of complex, robust systems will be another important research area for the CMI. To do this, scientists from different disciplines will have to come together, transcend their respective disciplines, and broaden the scope of their research.

CANDES: Absolutely, and at the same time we want to rethink computation, particularly large-scale computation. A trivial answer to the large-scale problems is: give me more flops. Here is an area where mathematics could play a role by providing a more efficient data structure through more efficient representations of operators for calculation.

There's another very interesting avenue that we will explore—while the world we live in is continuous, and we have the laws of physics formulated in a continuous way, computers are only able to handle equations and sets of data that are discrete and digitized. So if you're looking at numerical schemes, or if you digitize an equation, you have violated a lot of physical conservation laws that nature prefers to be preserved. How can you think really discrete all the way through without violating physical laws in your end results? That's a topic people will gravitate around, and that scientists at Caltech have already started attacking. Squarely addressing this challenge will be critical for moving beyond this limited, digitized computational view, to one that takes into account that the real world is continuous, multi-scale, dynamic, and complex.

PERONA: We hope the Center will bring the pure mathematicians at Caltech in contact with the technologists. We will be working very closely with the theorists in the physics center [CPI] as well.

EFFROS: Making that connection between pure mathematics and applied mathematics is critical. You would be amazed how broadly our theme sweeps. There are people in economics, humanities, and social sciences who are worrying about the mathematics of information. There are people all over campus who are thinking deeply about the mathematics of information. The goal in many senses is to bring them all together.

CANDES: I'd also like to emphasize that the CMI will provide a real link to and between the other ISTI centers. ISTI will bring the divisions of Caltech together in profound ways, and this particular Center will be the glue for ISTI.

PERONA: At the beginning, creating this Center felt like a construction. But now it feels like an inevitable fact. It seems impossible not to have thought about it a little bit earlier and it seems impossible that it will not exist. I see signs, all over the country, that the best, young creative people in every area that deals with information are just bursting out of the seams of existing fields. And this Center is going to capture them. We hope to attract the best talent in the country, both at the level of graduate students and at the level of young faculty. They will want to come to Caltech because this is the place where we destroy all the boundaries between disciplines and even the concept that the disciplines need to exist—we're focusing on the real problems of today. ENG

Emmanuel Candes is Assistant Professor of Applied and Computational Mathematics. Michelle Effros is Associate Professor of Electrical Engineering. Pietro Perona is Professor of Electrical Engineering and the Director of the National Science Foundation Center for Neuromorphic Systems Engineering at Caltech.