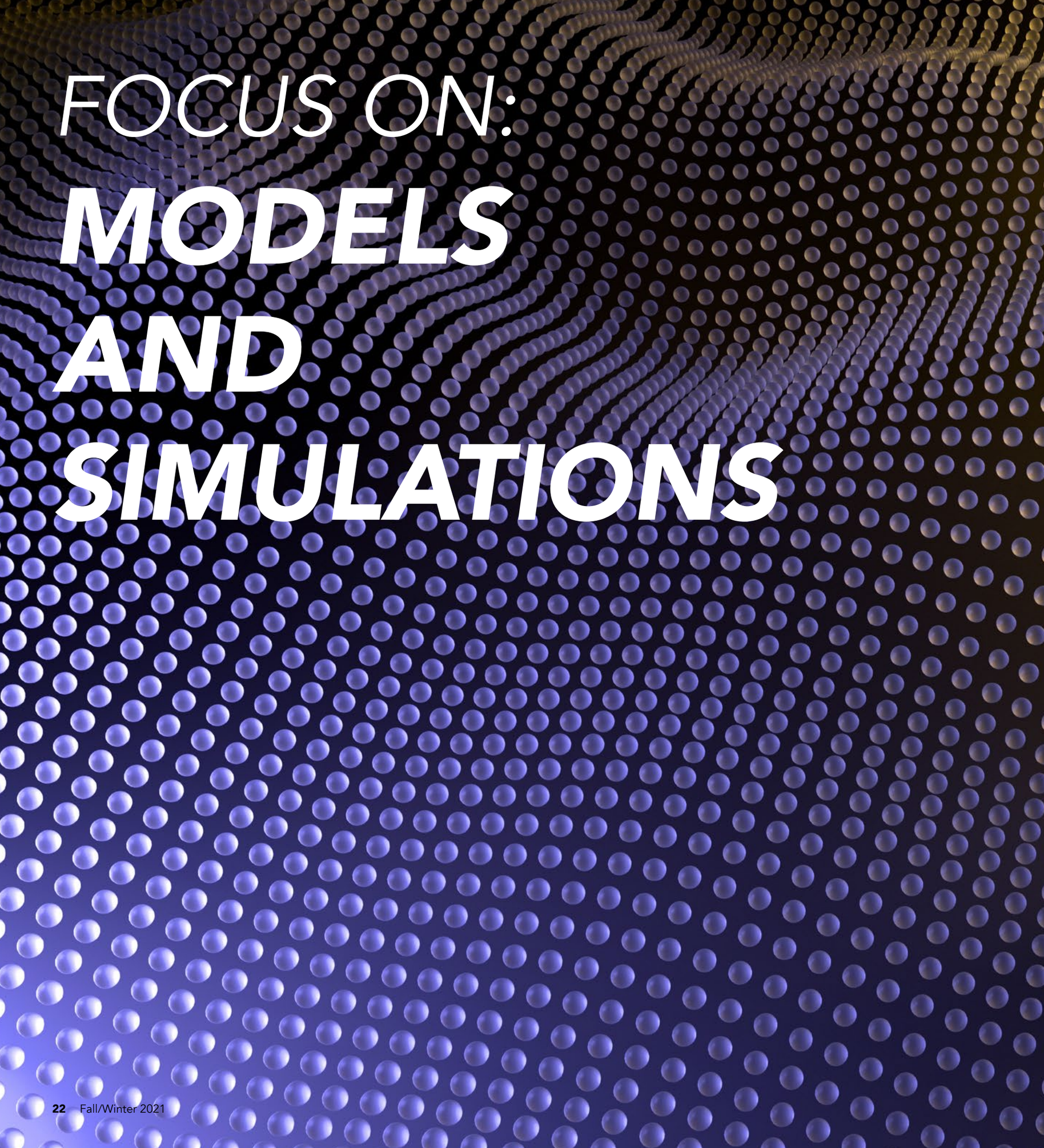
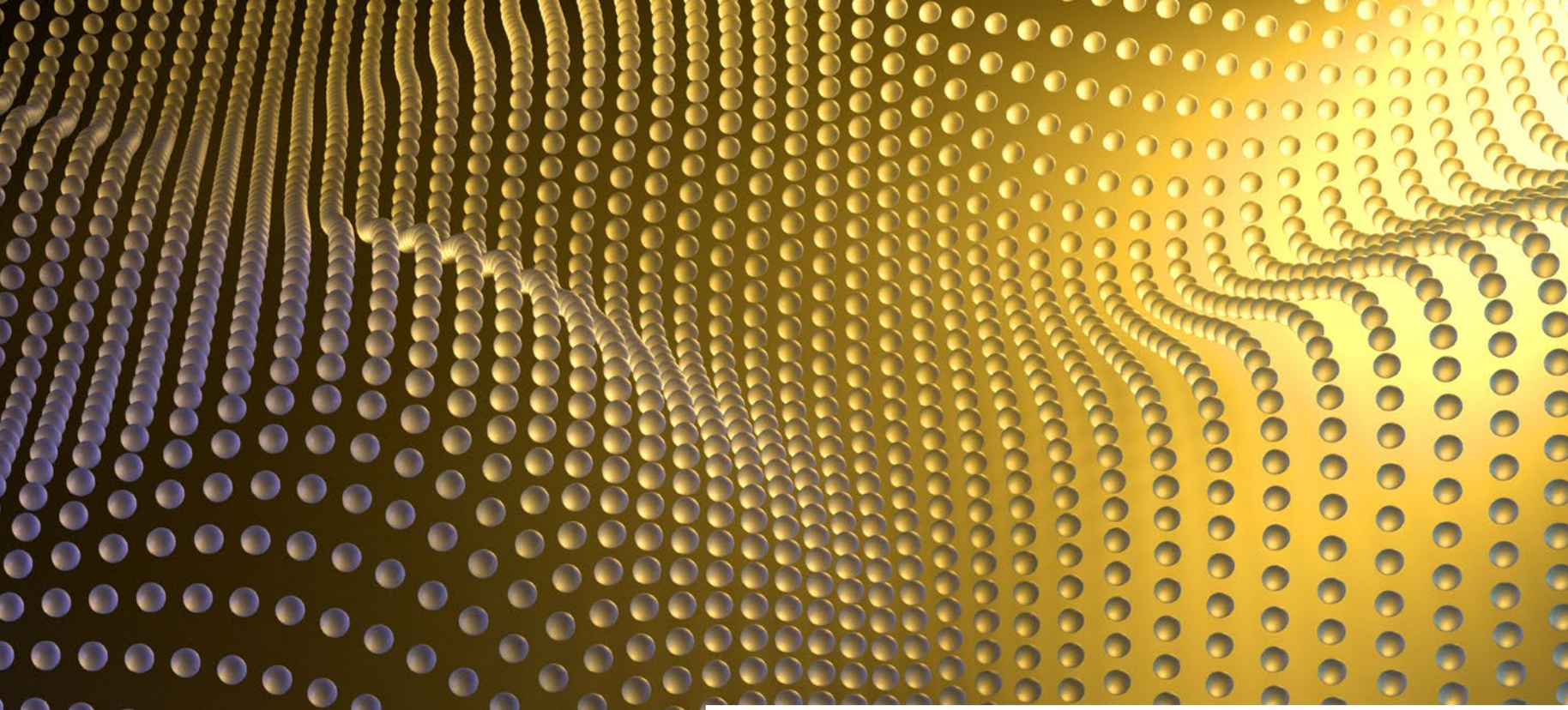


FOCUS ON: MODELS AND SIMULATIONS



Our coverage of research in *Convergence* often mentions theories, models, and simulations. But what exactly are those things, and how do they differ from or complement each other? Why are they used, how are they made, and what do they add to the research endeavor?

"I use the term *model* to refer to any approximate description of a material or a process," says materials professor **Chris Van de Walle.** "A theory provides a fundamental description, while a model is meant to create an *idealization* of real behavior to a given accuracy. I think of a simulation as the computer code that is developed and run to implement that model."

Fundamentally, says **Linda Petzold,** professor of mechanical engineering and computer science, "A model is simply a representation of something that is happening."

"My working definition of a model is 'any relationship that helps explain how a system, a device, or some aspect of nature behaves,'" says chemical engineering professor **Mike Doherty.** "That means the math equations you wrote down and the many assumptions that went into writing them."

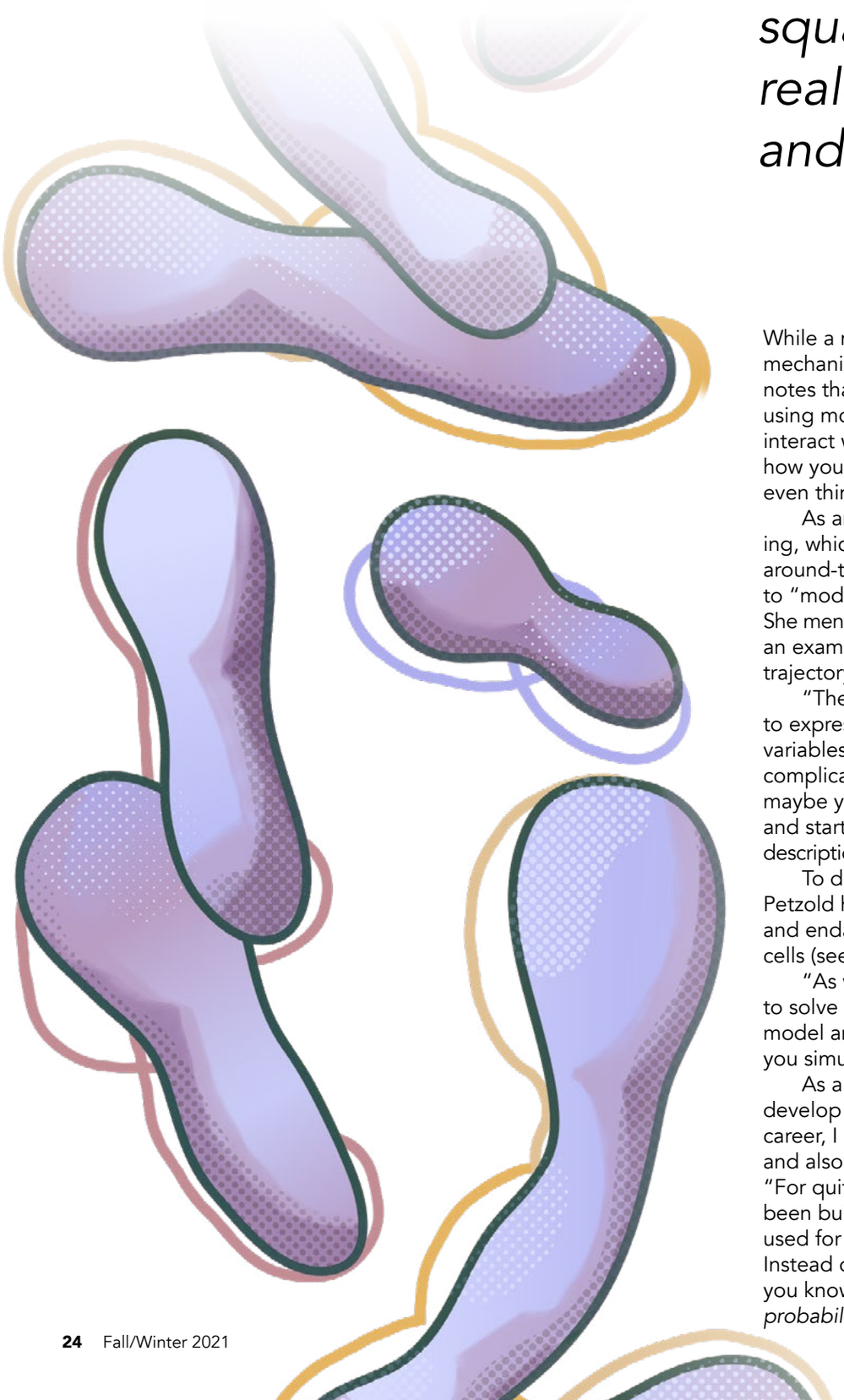
Chemical engineering assistant professor **Sho Takatori** says that his lab develops models "to describe various phenomena in physics, biology, and chemistry. A model is somehow representative of the world out there."

And to **William Smith,** professor in the Department of Molecular, Cellular and Developmental Biology, a model is "an experimentally tractable system that approximates a less tractable, but ultimately more significant, system."

In this issue's section, these five professors, plus postdoctoral researcher **Angela Zhang,** discuss theory, modeling, simulation, and their application in the context of specific projects in the professors' labs. Each offers rich insights, a unique perspective, and unbridled enthusiasm for the pursuit of discovery.

LINDA PETZOLD

Versatility lands her squarely in the twin realms of modeling and simulation



While a model can seem mysterious and esoteric, UC Santa Barbara mechanical engineering and computer science professor **Linda Petzold** notes that all of us are proficient model-makers: "You are making and using models all the time," she says. "You have a model of how you interact with different kinds of people, of how you drive your car, and how you get to the grocery store. You're so used to it that you don't even think about it."

As an applied mathematician, Petzold does mathematical modeling, which, she admits, is quite a bit more complex than the "walking-around-the-world" models she mentions above and allows researchers to "model all sorts of things by using mathematics to describe them." She mentions Newton's second law, F=MA (force = mass x acceleration), as an example of a mathematical model; it allows one to predict, say, the trajectory of a thrown a ball if its initial position and velocity are known.

"Then there is computational modeling, which can be used to express equations representing phenomena having many more variables," she continues. "In the example of throwing a ball, you might complicate the model by adding humidity, wind, or temperature, or maybe you want to know what happens to the ball if it falls into a lake and starts to sink. If you build enough information into the mathematical description, you can expect some predictive power."

To describe a complicated system or process with a model — and Petzold has modeled everything from automotive suspension systems and endangered frogs to migraine headaches and the mating of yeast cells (see below) — she first develops a system of differential equations.

"As we add more complexity," she says, "it becomes impossible to solve the equations exactly. At that point, you build an approximate model and use it to simulate the system on a computer. In other words, you simulate the model."

As a computer scientist, Petzold also works with her group to develop the computational software to run simulations. "Earlier in my career, I built software for simulation of differential equation systems, and also for differential equations subject to constraints," she explains. "For quite a while now, my research group and our collaborators have been building software for discreet stochastic simulation, which is used for systems or phenomena that are characterized by randomness. Instead of having something deterministic, like Newton's law, where you know that doing X will result in Y, stochastic simulation posits the *probability* that something will happen under a certain condition or

“Who knows why the yeast cell prefers that one over the other one? What motivates this? We wanted our model to replicate that kind of behavior. So, we started thinking, *What would it take for the model to be able to produce that?*”

set of conditions. This can involve a lot of computation.”

In recent years, Petzold has frequently modeled biological data at the cellular level, an area rife with complexity. “A cell is so complex that nobody can really look at it and say with certainty, ‘This is what’s going to happen next,’” she says.

One of her recent projects involved modeling the dynamics of yeast mating. “You probably haven’t thought much about yeast mating, but yeast can mate!” she quips. “They sense a pheromone from another yeast cell and grow a projection, called a *polarisome,* in that direction.”

How Petzold became involved in the project speaks to the often-serendipitous pathways that lead her to specific research. “It seems that every research project begins in a different way,” she says. “It can come from anywhere. It can come from someone asking, ‘Why doesn’t your software do this’ And I’ve thought, *Why would anyone want to do that?* And then I find out and think, *Well, that’s really interesting,* and then I might get pulled in that direction for a while. The initial impetus can just as easily come from the mathematics or science side of it, too.”

In the case of the yeast cells, the opportunity arose after she took a seat in the front row at a conference, following a talk she gave about her stochastic simulation software. She recalls, “The next speaker, Dr. Tau Mu Yi, then a faculty member at UC Irvine, was sitting next to me and asked me if I had any specific biological problems in mind for our spatial simulation software And I said, ‘No, and we really want to find one.’ And he said, ‘I think I have one.’ We talked about it, and that was the beginning of our yeast mating collaboration.”

Many biologists Petzold has worked with begin by describing the system of interest via a simple cartoon-style drawing having several panels, similar to a comic strip or an advertising story board. “But these things can get really complicated, and they can have feedback loops in them. There can be processes that have different time scales — fast ones and much slower ones — that all work together,” she says. “A mathematical model essentially quantifies the processes illustrated in the cartoon. As it develops, the model can become quite complicated, as there are usually a great many processes and interactions to keep track of, so we simulate it on a computer.”

For the yeast-mating model, she says, “We began with fairly simple models and added more processes, as needed, to explain the data. For example, you can begin with a model of the pheromone response, and from there you ask, ‘How does it form this structure, the polarisome?’ And then, there are observed properties that we wanted the model to mimic; for example, sometimes the cell makes a polarisome spontaneously, apparently without reason, with no mate nearby. We wanted the model to be able to replicate that behavior.”

If models sometimes don’t give the expected results, they can also give *partial* results, making it necessary to adjust them. “In the yeast project, we knew that it can also happen that two yeast cells are attracted to each other and start growing toward each other. Then, another potential mate arrives from another direction, and one yeast cell may start retracting the polarisome it had made initially and start growing it on the other side toward the new mate,” Petzold explains. “Who knows why the yeast cell prefers that one over the other one? What motivates this? We wanted our model to replicate that kind of behavior. So, we started thinking, *What would it take for the model to be able to produce that? And of all these possible unknown reactions, what could be the one that could make this happen without screwing up any of the other capabilities of the model*?

“It takes a team of people to do that; you need mathematicians, computer scientists, engineers, and biologists,” she continues. “Finally, we had a quite-plausible model that was supported by experimental data and could reproduce the growth of the polarisome in one and two dimensions.

“But when it came time to extend the results to three dimensions, we met with a surprise: the polarization was not stable — it drifted from location to location — which is not what happens in nature! We postulated, in collaboration with our then–UCSB mechanical engineering colleague **Professor Otger Campàs,** that in nature, stabilization of the polarisome might be achieved through a mechanical feedback. It was a new wrinkle, but it was correct, and after quite a bit of research, we elucidated this feedback and incorporated it into our three-dimensional model, which now responds like a yeast cell.”

MICHAEL DOHERTY

When models describe realities that are not yet known

Models often inform experimentation, but experimentation can also validate a model. UC Santa Barbara chemical engineering professor **Michael Doherty,** whose group is known for modeling and simulating chemical engineering processes, is intimately familiar with that terrain. "The most interesting models are those that predict something that no one has ever experienced or seen before," he says. "We've done this probably twice in my career, and we've used experiments specifically to test those models."

But to test a model, you first have to develop one, and before doing that, says Doherty, "The modeler has to decide which type of model is needed. Am I going to build a *deterministic* model, in which all of the quantities are typically averaged quantities, like temperature, expressed as an average of the kinetic energy of all the molecules in the system? Or do I need a *stochastic* model, which explicitly takes into account the random microscopic behavior of molecules in a gas or a liquid? I might make a *steady-state* model, one in which time has no part in how the model is formulated, or a *dynamic* model, in which case time is an intrinsic variable in the system. You need different classes of mathematics and different types of modeling schemes depending on your purpose and the nature of the system. It's a conscious decision."

Whatever type is needed, Doherty says, "You start with the simplest possible model you can. You strip down all of the phenomena to the absolute minimalist set needed. From there, you can build a higher-fidelity model if you feel the need to do it.

"You can make a qualitative model. For example, in the shower, you know qualitatively that if you turn up the flow of the hot water, the shower gets hotter. There's no mathematical model involved, just a qualitative cause-and-effect relationship. For many aspects of modeling, that works fine, but say you want to know exactly how much you should turn up the hot water to reach a certain temperature. Now, you need a mathematical model that does an energy balance and a mass balance and tells you what's required to increase the flow rate to get to the desired temperature. If you want a quantitative model, which you almost always do, you have to develop some mathematical equations and ultimately solve them in order to figure out what's going on."

THE FIRST BREAKTHROUGH

In his first major finding that was confirmed by experimentation, Doherty used modeling to predict the existence of a state called a *reactive azeotrope,* which, he says, "is very important in chemical separations but also something that people previously had no clue about," despite the fact that for two hundred years before it was discovered, that state had limited the effectiveness — and therefore the use — of a complex process in which a chemical separation and a chemical reaction are conducted simultaneously. Under certain circumstances, the separation process would continue for only so long, and then, suddenly, stop; no further separation was possible. "Nobody understood why," Doherty says.

His model posited that the reactive azeotrope existed and that it also stopped the process, and his group wrote papers about it, despite there being "not the slightest scrap of experimental evidence that this was in any way part of reality," he recalls. "It had never been predicted or anticipated, and while some phenomena arising as a result of it had been seen in experiments, no one could explain them."

Year after year, Doherty, now the Mellichamp Chair in Systems Engineering at UCSB, attended meetings where plenty of experimentalists were present. And he would suggest to them that if they would plug key quantities from their experiments into his equations, they would know whether they had found a reactive azeotropic state.

But nobody did, so rather than waiting, he and his colleague Michael Malone decided to do it themselves. "We scrounged up some funding and built the equipment and hired a really great experimentalist, Wei Song, now a technology leader at Shell," he recalls. "And sure enough, we showed experimentally — absolutely and conclusively — that this phenomenon is real. We discovered it in a real mixture with a real reaction and real separations and in 1997 got a very nice paper, titled "The Discovery of a Reactive Azeotrope," published in *NATURE*, which is not a place I publish as an engineer."

With experimental evidence to validate the model, interest grew, and Doherty and Malone, currently Vice Chancellor for Research and Engagement at the University of Massachusetts, Amherst, later wrote about the work in a broader context in their widely read, extensively cited book *Conceptual Design of Distillation Systems.*

> **"Frequently, you want a particular polymorph for a particular application, and the question is, how do you get the one you want and not some other one?"**

THE SECOND ACT

From there, Doherty switched tracks, teaching himself a whole new field, a process he likens to "getting another PhD," which led to his spending the past roughly twenty years focused on crystals. "We and others have devised ways of writing down mathematical models for how molecules are attached to crystal surfaces, and those models can tell us the rate at which they attach, which can tell us the speed at which a crystal face will grow," he explains. "We then perform computer simulations to solve the equations and determine the outcome, given the molecule, its environment, and so on."

Simulations and models are extremely close, highly interdependent, mutually indispensable siblings. "Just writing down a set of equations doesn't typically give me the slightest hint of what those equations are going to tell me about my system or phenomena or natural processes," Doherty says. "You actually have to solve the equations. If you can't, then you've learned nothing. That's where mathematics comes in. Simulation occurs when you have a model for which you cannot find a closed-form mathematical solution, which is most commonly the case, so you have to devise a numerical method that solves it. Typically, the models are so complicated that you could never solve them without a computer."

Recently, one of Doherty's students, **Thomas Farmer** (PhD '18), worked with him to make another important breakthrough, in crystal formation. "It was a model we put together that predicted a very exciting and unusual outcome that no one had discovered before and that could have a big impact on pharmaceutical manufacturing," he says,"perhaps one as profound as the discovery of the reactive azeotrope was in the world of separation and reactions." The discovery led the European Federation of Chemical Engineers to invite Doherty to give the 2021 Euro Danckwerts Memorial Lecture last September.

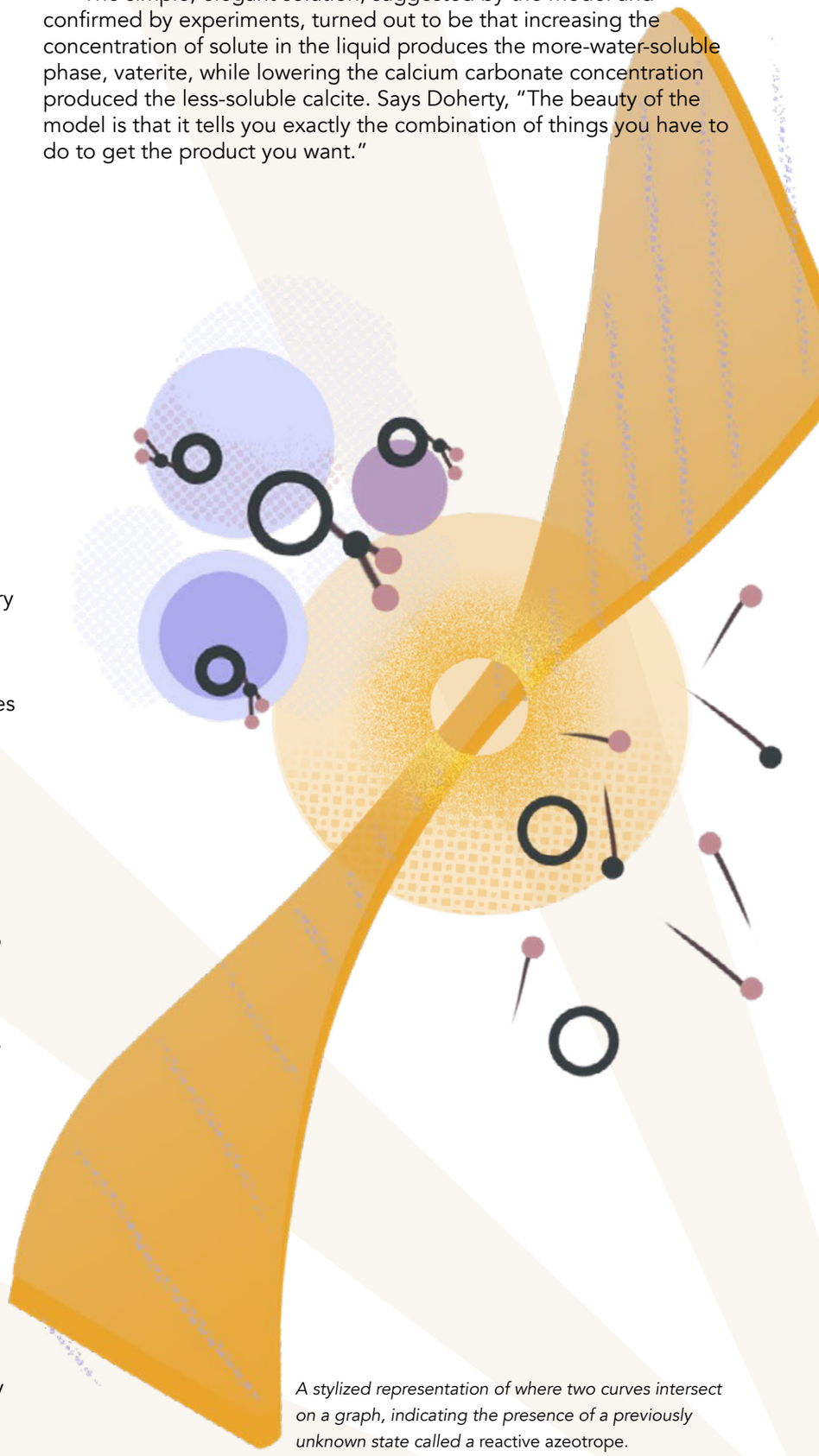
The work is based on the fact that most solids can crystallize in multiple different lattice configurations, or phases, and different phases of the same solid, each of which is called a *polymorph,* have wholly different qualities. The prime example is carbon, which, in its graphite form is soft, gray, opaque, and a good conductor of electricity, but that same carbon under enough pressure gets arranged differently to become diamond, a hard, transparent solid that does not conduct electricity.

"Frequently, you want a particular polymorph for a particular application," Doherty explains. "And the question is, how do you get the one you want and not some other one? We figured out how to do that and published a paper about the findings, which said, essentially, that if you design your system in *this* part of the space, you'll get this output, and if you design your system in *this* part of the design space, you'll get a completely different output. The two outcomes are mutually exclusive."

It is theoretically known and experimentally proven that the most stable polymorph is also the least soluble, and most chemical engineering processes naturally produce the most stable solid. "The problem is that the most stable form is also the least soluble in water, and we're made mostly of water," Doherty says. "So, while a lot of modern pharmaceuticals are effective, you can't get enough of them into you. The companies want a form of the drug that has much higher solubility in water. We were able to achieve better solubility on demand in a stable process that doesn't change or give unexpected, unwanted results even if you run the system for weeks."

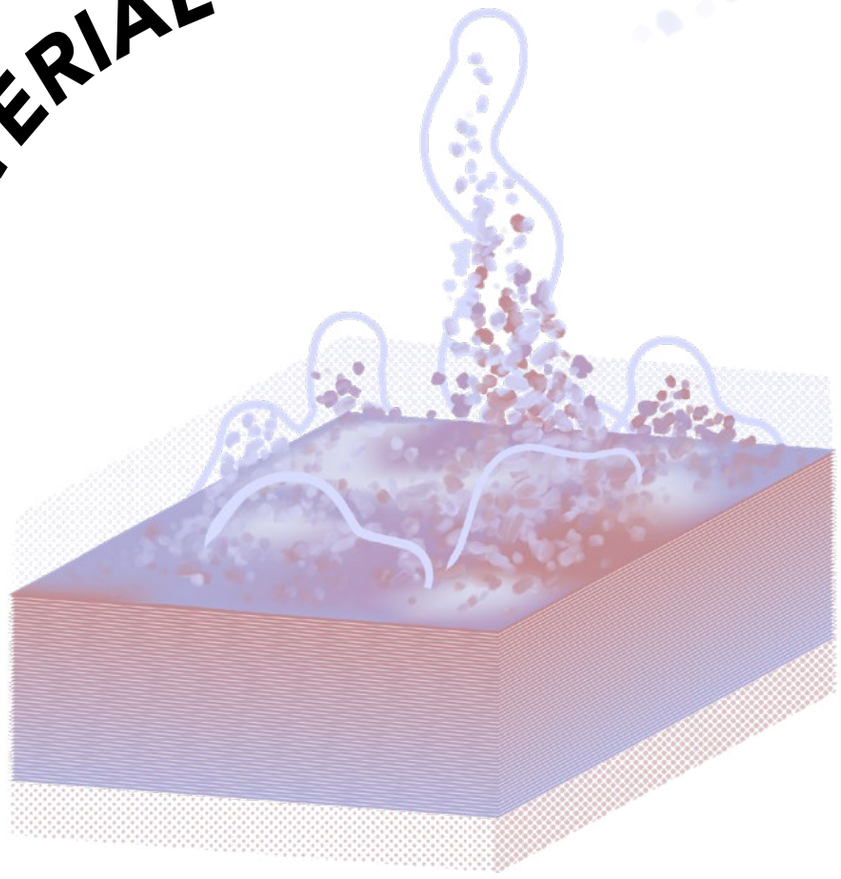
While the model was compelling, once again, experiments were needed to validate it, and this time Doherty decided not to wait. They built an apparatus and, over eighteen months, did the experiments to test the model. They determined how to crystallize their target material, calcium carbonate, into one of the two main forms it takes: calcite, which is the lowest-energy, the most stable, and the least water-soluble form; and vaterite, a very high-energy form, called *metastable,* that is more water-soluble.

The simple, elegant solution, suggested by the model and confirmed by experiments, turned out to be that increasing the concentration of solute in the liquid produces the more-water-soluble phase, vaterite, while lowering the calcium carbonate concentration produced the less-soluble calcite. Says Doherty, "The beauty of the model is that it tells you exactly the combination of things you have to do to get the product you want."



A stylized representation of where two curves intersect on a graph, indicating the presence of a previously unknown state called a reactive azeotrope.

MODELING BACTERIAL "PILE-UPS"

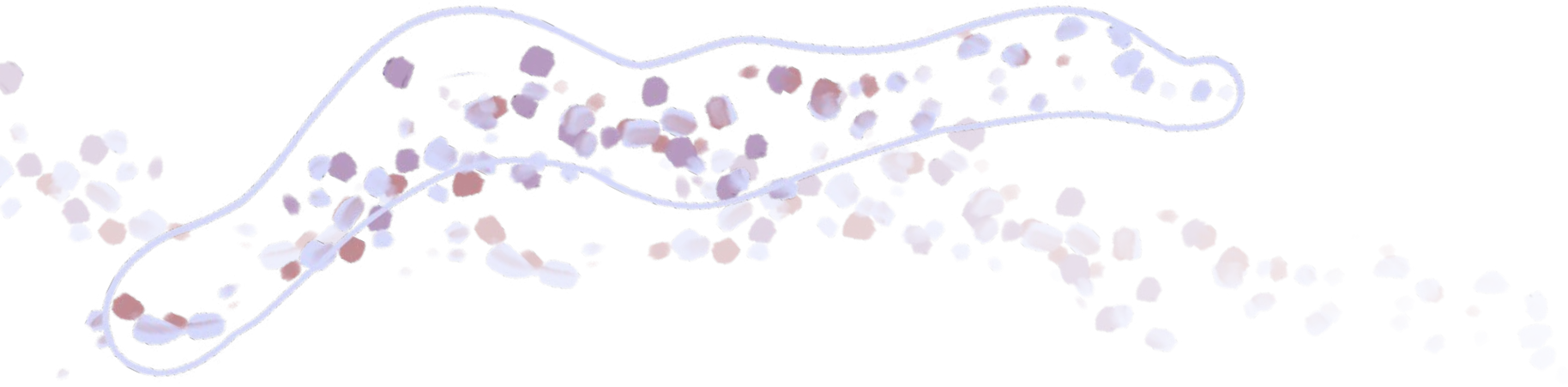


Sho Takatori tracks the swarming behavior of tiny organisms

Like many of us, UC Santa Barbara chemical engineering assistant professor **Sho Takatori** is fascinated by the swarming behavior exhibited by birds, fish, insects, and other organisms. A part of Takatori's research group studies such behavior, not in the sky or the ocean, but under an optical microscope and through models and simulations, to see how it plays out among populations of self-propelled particles such as bacteria. These particles exhibit autonomous motion and are classified in the category of *active matter.*

Such particles often show interesting dynamics, Takatori says, because they are able to communicate with each other as they move. "A fish swimming alone in the ocean might appear to be undulating in a random motion, but if you put it into a school of fish, you see this beautiful swarming pattern. Those patterns emerge because the fish can both move in a directed fashion and communicate with each other. Whenever you have a large collection of constituents that are able to make their own decisions individually, non-trivial patterns can come about. Our group tries

> **"It describes how a colony of bacteria initially organized as a densely packed, two-dimensional monolayer of cells can suddenly transform into a hierarchical three-dimensional stack of cells."**



to understand how that motion of an individual entity or constituent within a material might serve to dictate the overall properties of that system or material."

In one project, Takatori was interested in the swarming and aggregate behaviors of bacteria on surfaces. "A lot of times, they're adsorbed onto a surface, such as a biomedical device or other implant that is put into the body," he says. "We want to understand how colonies form and grow there and on other surfaces, such as soil, rock substrates in the ocean, or the surface of our lungs. Understanding the physics behind the swarming behaviors of such organisms is a rapidly growing field right now."

Projects in Takatori's lab are usually inspired, as he says, "by something really cool that happens in a simulation or an experiment or, if I'm doing a pen-and-paper analytical theory, and the model produces surprising results I would not have expected. I might wonder then if we could realize this experimentally or if a simulation would show that it is really happening. A lot of it is curiosity-driven."

In one recent project, Takatori's group, in collaboration with UC Berkeley chemical engineering professor Kranthi Mandadapu, examined a process that they refer to as *motility-induced buckling*. It describes how a colony of bacteria initially organized as a densely packed, two-dimensional monolayer of cells can suddenly transform into a hierarchical three-dimensional stack of cells. The project involved a combination of experiments, theory, and simulation in a way that required "a lot of careful modeling and understanding the different types of interactions that lead to such a transition," Takatori says.

Takatori sees multiple benefits arising from the varied competencies of his lab group. "Having the experimental side allows us to narrow down the very large parameter space of all the interesting phenomena that can happen and informs better theoretical models," he says. "Experimentalists can also tell us that something we are proposing in the model cannot be done experimentally. Models are useful at guiding the way by narrowing the pathways for experimentation, just as experiments are useful for refining the model."

In the motility study, the researchers began by performing experiments (on other projects, the equations and simulations might precede experiments), which showed that swarming motions exhibited by dense concentrations of bacteria moving around on a 2D surface could cause individual bacteria to pop out to a third dimension. "We observed this under the microscope and felt that it was an important process that could describe various sorts of other phenomena out in nature," Takatori says.

From there, they used a simulation to see if they could reproduce some of the results they observed in the experiment. "Our ultimate goal is to come up with a good pen-and-paper equation or an analytical model of this particular cell growth on a substrate," he says, adding, "If you have a good model, you don't need anything else, but often, we don't have the right, accurate model to begin with for very complex systems, so we need other tools, like computation and simulations, to help us develop better models for those systems."

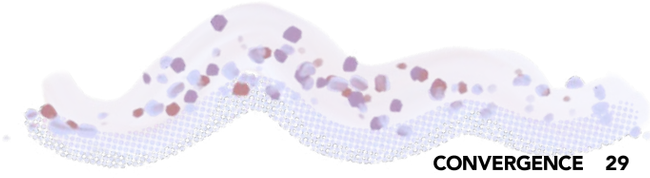
Even simulations can require further modeling. "We might need a model to describe the interactions among the constituents that are represented in the simulation," Takatori says. "For instance, we often do simulations of particles suspended in a fluid, and we need to model how the individual particles communicate with each other. Their communication is quite complex, so we need a model to describe it easily so that we can perform a large-scale simulation of it."

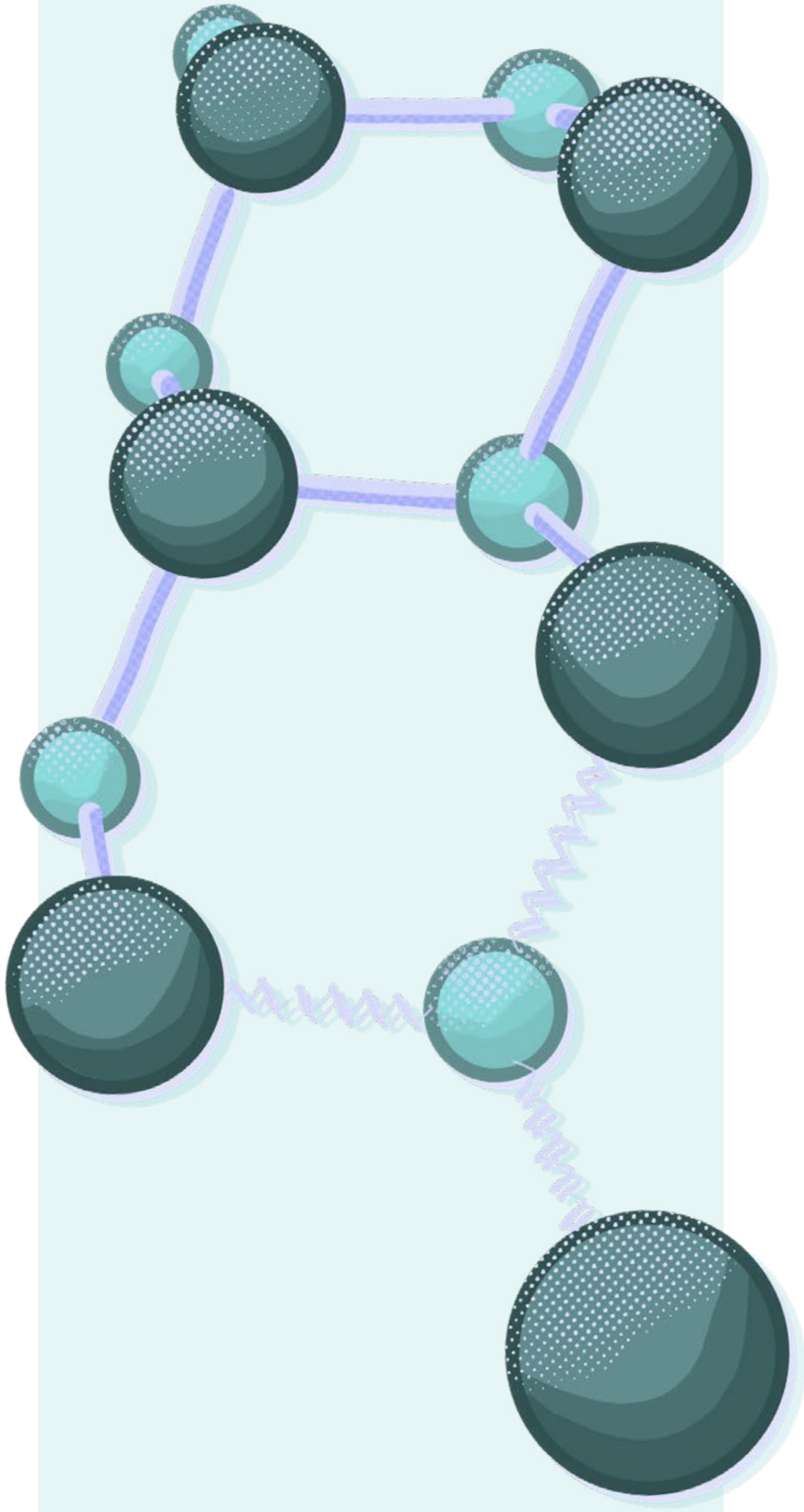
Takatori explains that researchers use different tools from the modeling-simulation-experimentation workflow, and use them in different ways, depending on the scales of time and length they are looking at. For those who study molecular systems, he says, "The simulation might evolve every single solvent molecule and every single hydrogen and carbon molecule. One usually performs a very short simulation, because one is interested in understanding the dynamics that occur on a very short time scale."

On the other hand, he says, "A lot of what our group studies takes place on much larger length and time scales. We're not looking at the motion of an individual water molecule. Rather, we're interested in suspensions of larger, micro-sized particles, like bacteria, suspended in a fluid. Because the bacteria are much larger than the solvent molecules, we can often do what's called *coarse-graining*, which means removing the solvent from the simulation, because the solvent molecules are much smaller than the particles we're interested in.

"If I wanted to do the full simulation that included all the solvent molecules, it would take longer than the academic career of a PhD student. We're focusing on the dynamics of the particles we're interested in as opposed to trying to evolve every single molecule and particle in the system. Computational tricks like coarse-graining can be used to accelerate the simulation dramatically."

He concludes: "Our goal, by combining theory, coarse-grained simulations, and experiments, is to advance a mechanistic understanding of complex fluid systems, such as bacterial films and particle suspension interfaces. Theory, simulations, and experiments each provide a different tool to study these systems across diverse length and time scales."





CHRIS VAN DE WALLE

Modeling in the quantum realm

In 2007, the laboratory group of UC Santa Barbara materials professor **Chris Van de Walle** began working to model *Auger recombination,* a key process in the functioning of semiconductor-based light-emitting diodes (LEDs). Their work was motivated by experimental reports showing a rapid loss of efficiency in LEDs when the current into them reached a certain level. Since efficiency is a key strength of the tiny, long-lasting lights, breaking through this high-current efficiency "droop" was driving research around the world.

Van de Walle explains the process: "In an LED or a laser, we inject electrons and holes — the positively charged counterpart of electrons — and when they meet, they annihilate, producing energy that is emitted in the form of light, a process called *radiative recombination*. But sometimes, when an electron and a hole recombine, no light is emitted; instead, the energy is released in some other way, a so-called *nonradiative* recombination process — Auger recombination — which creates a source of loss, or reduction of efficiency, in the device. Using first-principles calculations, we can calculate the rates at which these processes happen; and these rates then can be fed into a model that predicts the efficiency of a device.

"In principle, the findings about efficiency loss could be explained by Auger recombination, but the majority of the community did not believe this to be true," Van de Walle continues. "Their belief was based on older models that did not take into account vibrations of the crystal lattice of the semiconductor. We thought the vibrations were important."

Making choices is an important part of developing a model, Van de Walle says, and in modeling Auger recombination, his group made a choice — to account for lattice vibration — a choice that would have a profound impact on the entire semiconductor field. It paid off, leading them to discover that, as suspected, in some materials, such as gallium nitride, which is the key material for solid-state light emitters, and the one UCSB Nobel Laureate **Shuji Nakamura** used to invent the world-changing blue LED, "The vibrations make a huge difference. They were the source of a stubborn barrier to efficiency gains at high current!"

The model, developed over a three-year period, provided an abject lesson in the tradeoffs involved in modeling and simulation. "Including the vibrations makes the model a lot more complicated and makes the simulations a lot more expensive [because they require much more computer time to run]," Van de Walle says. "We went through the considerable effort to include the vibrations, and found that the results could beautifully explain the experimentally observed efficiency loss."

Subsequently, he notes, [UCSB materials professors] **Jim Speck** and **Claude Weisbuch** performed detailed, highly original experiments in which they actually observed the "Auger electrons" — electrons that are kicked up to higher energies during Auger recombination. "That clinched it!" Van de Walle recalls. "The findings changed the perception of the role of Auger recombination in gallium nitride.

"Now that we have the model, we can apply it to other materials," he adds. "It has produced very useful insights for materials ranging from lead selenide, used in infrared detectors, to hybrid perovskites, used for novel solar cells. We discovered that Auger recombination in hybrid perovskites is greatly enhanced by a particular feature in the band structure of the material, and that this feature can be tuned my substituting some elements for others in the crystal, for instance, replacing a fraction of the lead atoms with tin atoms. Even in the absence of full-scale simulations, this model can tell the people who synthesize the material how to optimize it!"

There is a good deal of back and forth between the model and simulations, Van de Walle says: "Once the model is developed, we need to run the simulations to generate the data that the model needs. Even though we use established codes [written by groups that are experts at developing them] to solve the key equations, additional software needs to be written to process the output of those simulations for specific materials and calculate the quantities needed for the model.

"Along the way, we need to balance accuracy — and determine what level of accuracy is required — with computational cost; it's always a tradeoff. But once we have demonstrated that the simulations are 'converged,' or have met the desired tolerance requirement, we can produce results that elucidate the physical mechanisms. This, in turn, makes it possible to redesign devices to avoid the efficiency bottlenecks."

And that is where fundamental, first-principles science meets the marketplace.

Van de Walle's group works primarily at the scale of atoms, performing simulations of materials based on fundamental quantum-mechanical calculations. The output of those simulations can be used to do various types of modeling, such as the one his group developed to describe Auger recombination.

"Many beginning students in the field have a tendency to just run a lot of simulations, hoping that somehow a result will jump out. It never does. Before running any code, you need to carefully think through the problem and develop a model that will allow you to test hypotheses."

The mathematics are particularly complex, because, as Van de Walle says, "To model materials at a fundamental level requires very complicated equations that describe the behavior of electrons in materials. That behavior is governed by quantum mechanics — essentially a complicated version of the Schrödinger equation that was proposed almost a century ago to describe particles at the atomic scale. The solutions to that equation allow us to predict almost any property of any material."

Modeling, Van de Walle says, "allows us to develop and test hypotheses about materials behavior, thus guiding experimental growth and characterization efforts." And the first step in developing a model, he notes, is always the same: to think — slowly, deeply, thoroughly — about the problem to be solved. "Many beginning students in this field have a tendency to just run a lot of simulations, hoping that somehow a result will jump out," he says. "It never does. Before running any code, you need to carefully think through the problem and develop a model that will allow you to test hypotheses."

Once a model is developed, Van de Walle adds, "The most important thing is to verify and validate. Verifying means ensuring that the model actually returns what is desired. Validating means comparing the predictions with experimental results for a well-established case. Such benchmarking should also include checking the model's sensitivity to input parameters and establishing over what range of conditions the model is valid. Running computational simulations puts meat on the bones of the models, which is essential for validating the model and for making direct connections with experimentation."

Simulations can serve various purposes. "In some cases, we aim to characterize behavior or performance," says Van de Walle. "For instance, when experimental results are puzzling, simulations can help to provide understanding, as they did in the case of Auger recombination. In other situations, we make predictions that might guide the development of novel materials, or new combinations of materials."

Asked if he finds reward in modeling, Van de Walle does not hesitate. "Absolutely!" he says. "As an engineer, I like solving puzzles, and I like having an impact on how technology develops. It's very gratifying to see the papers that report our key insights being heavily cited. It means that other scientists are actually using those results and benefiting from them."

WILLIAM SMITH

The Ciona, Model Organism

William Smith, a professor in the UC Santa Barbara Department of Molecular, Cellular and Developmental Biology, understands as well as anyone that mapping the neural circuitry of the human brain and tracking the coordinated interactions of our roughly 85 *billion* neurons is a daunting task.

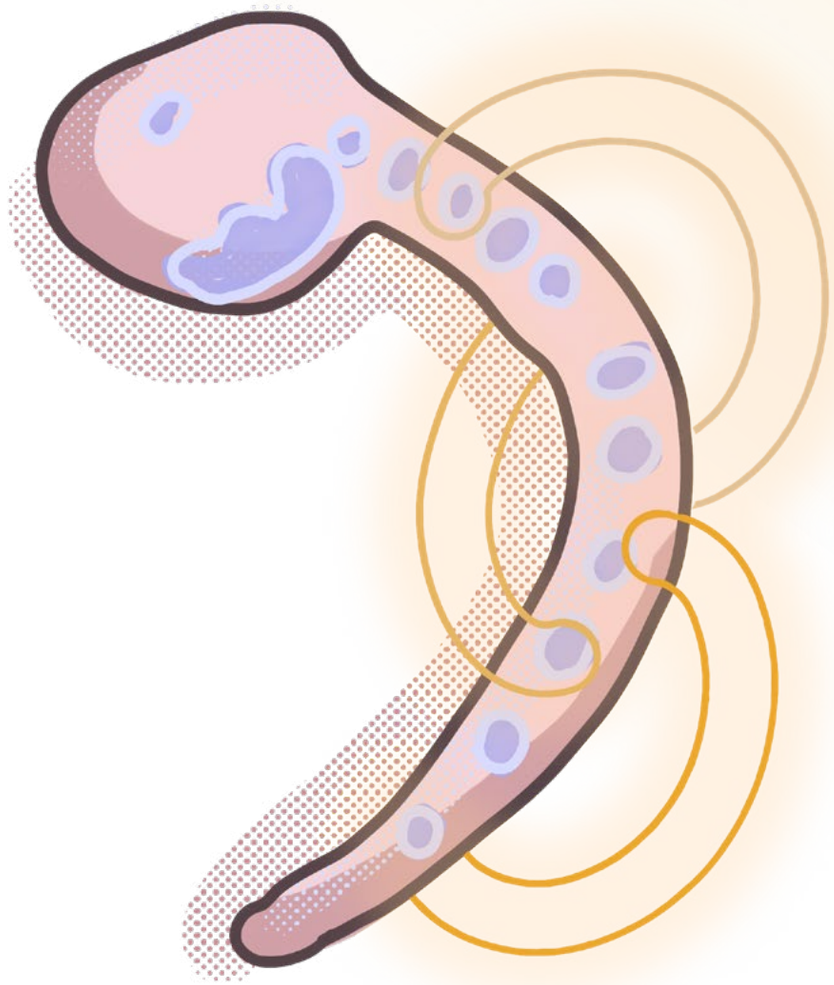
"Ultimately, we'd like to understand how vertebrates, animals like us, work," says Smith, whose research is focused on the nervous system of animals in the chordate phylum of animals, which includes vertebrates. "But we're enormously complicated, and at this point we're very far from a full understanding of our complex nervous system at the individual neuron level; there are just too many neurons."

A better place to start, he thought, would be to model the neural circuitry of a simpler chordate organism, to serve as a surrogate for — a kind of model of — the human neural system. Enter the ocean-dwelling *Ciona*, a primitive chordate that is closely related to vertebrates but has only 177 neurons. As one of few universities to have marine science facilities on its main campus, UCSB is ideally situated for this research.

The work, funded by NIH's Institute of Neurological Disorders & Stroke, began with Smith's collaborators at Dalhousie University in Halifax, Nova Scotia, gathering electron microscopy (EM) images of every neuron in a Ciona larva and then creating a wiring diagram, or *connectome*. Even with so few neurons, the project took many years of work by the Dalhousie team. Their results can be seen as an interaction matrix showing the 6,618 synaptic connections among the 177 neurons. "To do that for the human brain would require a matrix 2.17×10^{17} — or a billion trillion — times bigger than that," Smith says.

Beyond mapping the *Ciona*'s neural circuitry, the researchers wanted to determine which of six key neurotransmitters are present in which neurons. "It's important to know, because that tells us how it will act on its targets. Neurons can be excitatory — they can tell other neurons to have an action potential downstream — or they can be inhibitory," Smith says. "It's not a trivial undertaking."

To begin that work, Smith teamed up with UCSB Electrical and Computer Engineering professor **B. S. Manjunath**, Co-PI on the NIH project and an expert in computer vision and machine learning, and his student **Angela Zhang,** who received her PhD in September 2021, in an interdisciplinary undertaking of the kind that is much valued at UCSB. Zhang began by developing a method for overlaying the EM



data onto the light microscopy (LM) data from Smith's lab, bringing the two kinds of information into alignment, or *register,* so that specific organs in the Ciona, seen with LM, could be correlated with specific neurons identified in the EM data.

The light microscopy was conducted using *in situ hybridization,* which involves incubating fixed *Ciona* larvae with fluorescently tagged antisense RNA probes. Each RNA probe differs depending on the type of neurotransmitter marker it is meant to find, and when the probe locates the target, it fluoresces in a color correlated to that neurotransmitter.

The light microscopy allowed Smith's group to do a three-dimensional reconstruction of the actual nervous system, but not at synaptic resolution, which requires electron microscopy. "We can see the outlines of cells here, but not all the little fine synapses," he says, pointing to a photograph of an LM-generated image.

"Maybe we know the nucleus location of a cell in both sets of data. We have the shape information of this cluster of cells from the EM and of another cluster of cells from the light microscopy," Zhang says. "I have to use computation to rotate and stretch and kind of move the cells in one set of data so that it overlays well with the other set. If we get the alignment right, we can see how a cell in the EM data corresponds to one in the LM data."

The registered data can then be correlated with the location of neurons to deduce which neurotransmitters might be present at specific synapses. "We got pretty good matches for certain parts of the brain, such as the photoreceptors," Zhang says. "We could match the location of the neurons in the two sets of data."

Smith's team could then use logic and their own eyes to go further, for example, identifying specific neuron types. "We know from the EM data that there is a cluster of cells located by itself, between the head and the tail, and that it is emitting only one type of neurotransmitter, GABA [the inhibitory neurotransmitter gamma-aminobutyric acid], for example, because the probe is lighting up with the color associated only with GABA."

Things were less resolved in other locations. "They'd see something that looked kind of like a hand and fluoresced in many different colors," Zhang explains. "They knew the location of the cell and its nucleus from the EM data, but the intermingled colors of fluorescence prevented them from knowing what was going on. But, by combining many samples, we were able to make predictions with high confidence for most neurons."

Smith says that by allowing his team to model what kinds of signals are sent between neurons, Zhang's initial work "led to some new and interesting models about how the nervous system in this animal works," for example, that many of the behaviors in *Ciona* are evoked by disinhibition, or one inhibitory neurotransmitter inhibiting the (inhibitory) influence of another inhibitory transmitter, thereby allowing an action to occur.

After converging the LM and EM data, he adds, "Angela would come up with a hypothesis, saying maybe, 'I think these cells are inhibitory and connect with other cells that also appear to be inhibitory,' and then we'd do various behavioral and pharmacological experiments that would allow us to say, 'Yes, that really is how it works.' It's modeling, where we're working together."

Next, Zhang built a deep-learning model to try to complete the work of correlating specific cells with their neurotransmitter types. The in situ hybridization had made it possible to identify with confidence some neurons and the various neurotransmitters they make. She used those confirmed data points to train the model, providing it with examples of known excitatory and inhibitory synapses. She thought, *What if we used the EM data to look at the individual synapses and see if there are physical differences in their structures that can indicate whether it is an excitatory or inhibitory synapse?*

One challenge, Zhang says, was that, unlike many engineering problems, where the final answer is known and the task is to train an ML model to get the right answer, "With science, you don't actually know what the truth is; you're trying to figure it out."

The result, she says, is that "The model is able to learn, but we don't know what it is learning. Is it learning that a particular set of photoreceptors has a certain quirk, so it groups them together? Or is it learning something universal to all excitatory *Ciona* synapses? And can we trust the results of the model? We can't know now without more experiments, and I'm definitely not an advocate of trusting models blindly."

As Smith likes to say, "All models are wrong. Some are useful." Time will tell.

"Angela would come up with a hypothesis, saying maybe, 'I think these cells are inhibitory and connect with other cells that appear to be inhibitory,' and then we'd do various behavioral and pharmacological experiments and be able to say, 'Yes, that really is how it works.' It's modeling, and we're working together."