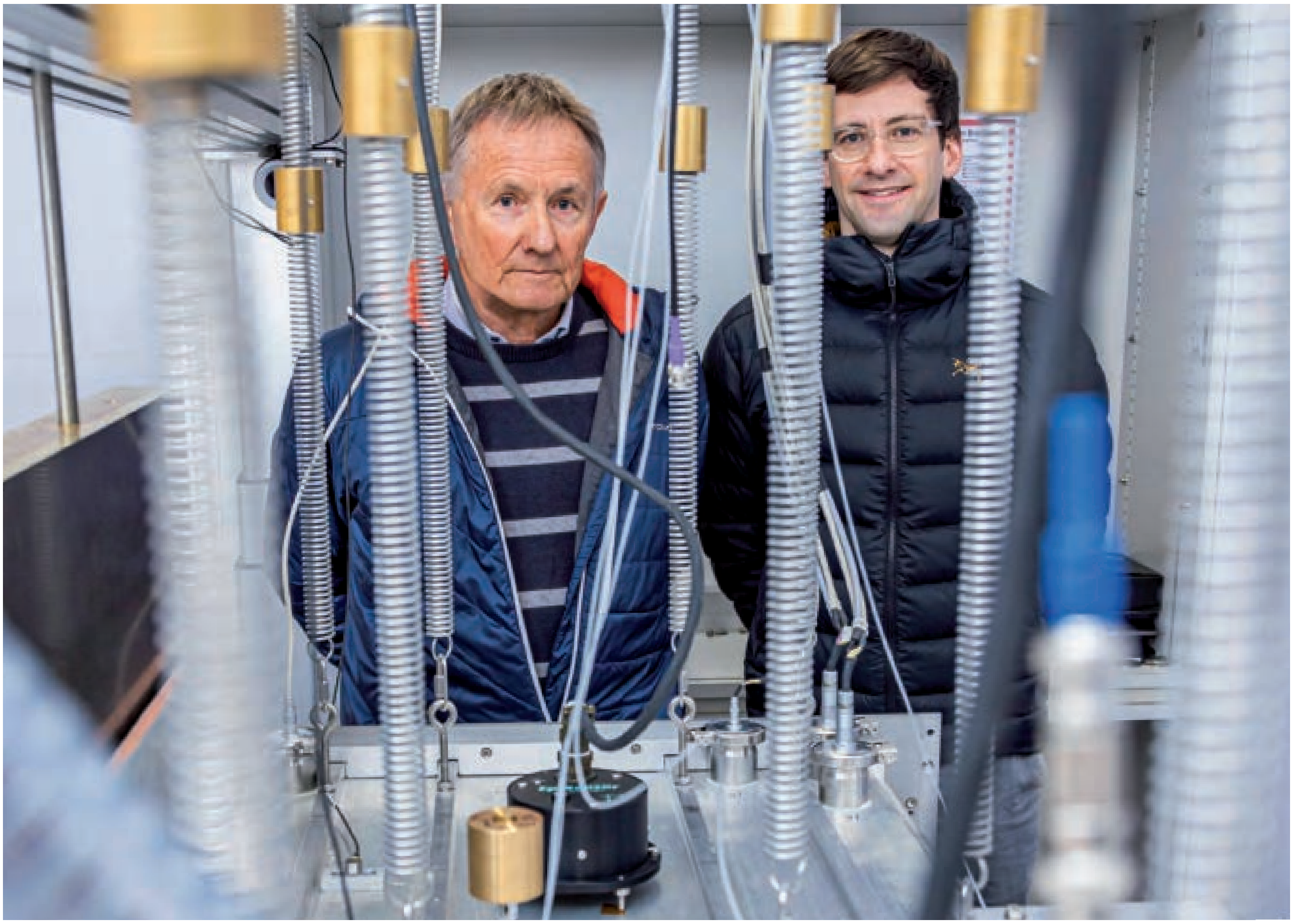
PHYSICS TODAY
July 2023 • volume 76, number 7
A publication of the American Institute of Physics

OCEAN OPTICS

Of insect swarms and magnets

The rise of free textbooks

A bottleneck for CO_2 storage




MFLI
MFLI
MFLI

SR830 ***Lock-In Amplifier***

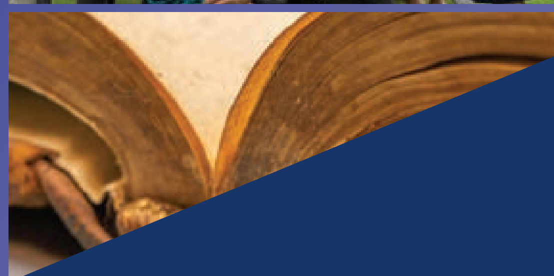
... the industry standard

SR830 ... $5250 (U.S. list)

- 1 mHz to 100 kHz range
- 100 dB dynamic reserve
- 5 ppm stability
- Auto gain, phase and reserve
- Harmonic detection (2F, 3F, ... nF)
- Time constants from 10 µs to 30 ks (6, 12, 18, 24 dB/oct rolloff)
- GPIB and RS-232 interfaces

The SR830 is the most widely used lock-in amplifier in the world. It has impressive analog specifications, many useful features, and it's reasonably priced. With over 100 dB of dynamic reserve and 0.01 degree phase resolution, the SR830 will handle the most demanding synchronous detection applications. It comes with convenient auto measurement features, a wide selection of time constants, and a built-in sine source.

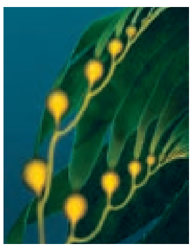
Everything you need in a lock-in amplifier!



PHYSICS TODAY

July 2023 | volume 76 number 7

FEATURES



ON THE COVER: This underwater photo, taken off the coast of California, shows giant kelp whose air bladders help it float in the ocean. Kelp is just one example of the many types of algae, both macroscopic and microscopic, that play a crucial role in ocean ecosystems. On **page 26**, Chuanmin Hu discusses the technologies that are helping scientists to better understand algae on a global scale and to make new discoveries in the field of ocean optics. (Photo by iStock.com/joebelanger.)

Recently on PHYSICS TODAY ONLINE

www.physicstoday.org

NASA/ESA/STSCI

Hubble tension

The two leading techniques for determining the universe's expansion rate, known as the Hubble constant, continue to deliver different values. Now a result from the analysis of multiple images of a distant, strongly gravitationally lensed supernova has been added to the mix.
physicstoday.org/Jul2023a

SMITHSONIAN

Continental crust

Earth's continental crust has lower iron content than does oceanic crust, and much of the iron takes the form of oxidized Fe^{3+}. A new experiment that simulates the conditions of magma deposits is helping researchers explore how iron gets depleted and oxidized in the molten rock that yields continental crust.
physicstoday.org/Jul2023b

NASA

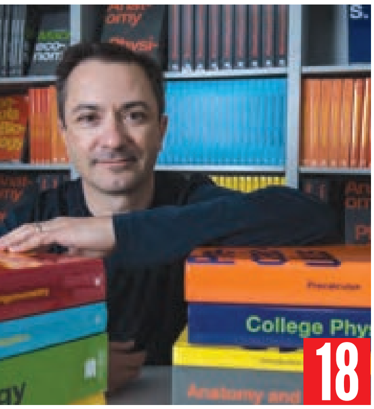
Saturn's young rings

Planetary scientists have long debated the age of Saturn's rings. In a recent study, researchers explored how long the rings have been collecting interplanetary dust. Using data from NASA's *Cassini* probe, which orbited Saturn for 13 years, the scientists estimate that the rings are far younger than the planet itself.
physicstoday.org/Jul2023c

PHYSICS TODAY (ISSN 0031-9228, coden PHTOAD) volume 76, number 7. Published monthly by the American Institute of Physics, 1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300. Periodicals postage paid at Huntington Station, NY, and at additional mailing offices. POSTMASTER: Send address changes to PHYSICS TODAY, American Institute of Physics, 1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300. Views expressed in PHYSICS TODAY and on its website are those of the authors and not necessarily those of AIP or any of its member societies.

PHYSICS TODAY

www.physicstoday.org



DEPARTMENTS

Editor-in-chief
Richard J. Fitzgerald *rjf@aip.org*

Art and production
Freddie A. Pagani, art director
Cynthia B. Cummings, photographer
Nathan Cromer

Editors
Ryan Dahn *rdahn@aip.org*
Toni Feder *tf@aip.org*
Abby Hunt *ahunt@aip.org*
David Kramer *dk@aip.org*
Alex Lopatka *alopatka@aip.org*
Johanna L. Miller *jlm@aip.org*
Gayle G. Parraway *ggp@aip.org*
Jennifer Sieben *jsieben@aip.org*
R. Mark Wilson *rmw@aip.org*

Online
Andrew Grant, editor *agrant@aip.org*
Greg Stasiewicz *gls@aip.org*

Assistant editor
Cynthia B. Cummings

Editorial assistant
Tonya Gary

Contributing editors
Paul K. Guinnessy
Andreas Mandelis

Sales and marketing
Christina Unger Ramos, director *cunger@aip.org*
Unique Carter
Krystal Amaya
Skye Haynes

Address
American Center for Physics
One Physics Ellipse
College Park, MD 20740-3842
+1 301 209-3100
pteditors@aip.org

PhysicsToday **@physicstoday**

AIP

Member societies
ACA: The Structural Science Society
Acoustical Society of America
American Association of Physicists in Medicine
American Association of Physics Teachers
American Astronomical Society
American Meteorological Society
American Physical Society
AVS: Science & Technology of Materials, Interfaces, and Processing
Optica (formerly The Optical Society)
The Society of Rheology

Other member organizations
Sigma Pi Sigma Physics and Astronomy Honor Society
Society of Physics Students

The American Institute of Physics is a federation of scientific societies in the physical sciences, representing scientists, engineers, educators, and students. AIP offers authoritative information, services, and expertise in physics education and student programs, science communication, government relations, career services, statistical research in physics employment and education, industrial outreach, and history of the physical sciences. AIP publishes Physics Today and is also home to the Society of Physics Students and to the Niels Bohr Library and Archives. AIP owns AIP Publishing, a scholarly publisher in the physical and related sciences.

SUBSCRIPTION QUESTIONS? +1 800 344-6902 | +1 516 576-2270 | **ptsubs@aip.org**

Better pay for grad students

Graduate employees at the University of Illinois at Urbana-Champaign rally in 2018. (Courtesy of the Graduate Employees' Organization at UIUC.)

News articles published in *Nature* and *Science* last year bring attention to a simple fact: Graduate student wages are low.[1,2] Data on median stipends for physics graduate students from the American Institute of Physics Statistical Research Center (AIP is the publisher of PHYSICS TODAY) confirm what we all know to be the case—that the field of physics is no exception.[3] At this year's American Physical Society (APS) March Meeting, a group of physics graduate students held an invited session on that topic.

An excellent piece by Jacqueline Acres in *APS News* last November highlights the ills that low pay levies on individual graduate students and the field of physics at large.[4] Some in physics view living through the low wages of graduate school as a rite of passage, something to weed out those who are not truly interested in the science—those who are not willing to subsist on ramen noodles and free seminar lunches for four to seven years in pursuit of their PhD. But that stoic view is wrong. Low graduate wages act as a barrier all right, but rather than keeping out those students who do not have the purest penchant for research, low wages keep out those from less privileged, less wealthy, marginalized backgrounds. Physics has always had a problem with diversity; raising graduate wages may help.

Low graduate wages are a national problem, and it is a problem that affects almost every field in academia. Within STEM (science, technology, engineering, and mathematics) fields, physics seems to be falling behind. At my institution, the University of Illinois at Urbana-Champaign, which has one of the largest physics graduate programs in the country, physics was for many years one of the lowest paid departments in the engineering college, rivaling the engineering administration program for the bottom spot. Through the efforts of the graduate labor union, a small group of concerned graduate physicists, and a supportive graduate director over several years, physics graduate students have gotten a raise. But the process was far from guar-

> "Some in physics view living through the low wages of graduate school as a rite of passage, something to weed out those who are not truly interested in the science—those who are not willing to subsist on ramen noodles and free seminar lunches for four to seven years in pursuit of their PhD. But that stoic view is wrong."

A student protests at the 2020 May Day Car Rally held by the Graduate Employees' Organization at the University of Illinois at Urbana-Champaign. (Photo by Ben Joseph Lash, courtesy of the Graduate Employees' Organization at UIUC.)

anteed, and record inflation last year has ensured that we are still paid far below the living wage in our city.[5]

What can be done? A national problem requires a national solution. Ad hoc raises to graduate wages on the level of individual universities are a Band-Aid fix (and, like Band-Aids, should be assessed and reapplied often while in use). The optimal solution requires national legislation. Physics organizations, such as APS and AIP, ought to use their government affairs and public policy offices to lobby Congress on the issue. It will be a long road, and only one of many issues addressed by such offices, but it should at least make the docket.

Graduate students in other disciplines are rallying around the issue. The biological sciences seem to be leading the way.[1] Physicists should be wary not to fall behind.

References

1. C. Woolston, *Nature* **605**, 775 (2022).
2. K. Langin, *Science* **376**, 1033 (2022).
3. American Institute of Physics Statistical Research Center, *Graduate School Support: Types of Support and Median Stipends for New Physics Bachelors Enrolled in a PhD Program*, Physics Trends flyer (fall 2022).
4. J. Acres, "The back page: Graduate students should be paid living wages," *APS News* (10 November 2022).
5. K. Shinbrough, J. Acres, "Physics graduate student pay data," American Physical Society Forum on Graduate Student Affairs.

Kai Shinbrough
(physics.grad.pay.campaign@gmail.com)
University of Illinois at Urbana-Champaign

"Quantum physics" doesn't make the Sun shine

I appreciate the work done by science communicators, as I imagine most other scientists do, and I believe their work is central to our goal of advancing humanity's knowledge of the world. My respect for science communicators, however, would be deeper if they used language more precisely in their descriptions of things. There's too much loose scientific language in the world today, which gives opportunists the chance to rebrand findings as snake oil.

Words can lose their meaning when science communicators pander to colloquial language. A frequent example in recent years is the use of phrases like "because of quantum physics." Such a phrase is confusing because the occurrence of physical events is not due to a body of scientific theories.

Here are some of the hits Google returned when I searched the phrase in quotation marks (to show exact matches):

> "The Sun only shines because of quantum physics."
>
> —*Forbes*

> "Quantum materials behave in surprising ways because of quantum physics."
>
> —Energy.gov

> "The bread toast which you enjoy while sipping on your morning tea is able to make its way to your plate only because of Quantum Physics."
>
> —StudiousGuy

> "Bubbles give off weird light when popped because of quantum physics."
>
> —*New Scientist*

> "Fireworks are only possible because of quantum physics."
>
> —Big Think

To be clear, only a physical mechanism can be the cause of a physical event. Not math. Not a theory. Fireworks have existed for well over a millennium—quantum physics for only a century.

If you say something physical happened "because of quantum physics," it's like saying it happened because of your illustrious Uncle Buck. Instead, state plainly what it is your Uncle Buck said about the physical mechanism that causes the physical event. Doing so will help you gain more respect as a science communicator.

Think of quantum physics as a salad. If you want to provide an explanation for a physical mechanism, don't just point to the whole plate—instead, tell your audience what's at work, whether it's the color of the cherry tomato, the shape of the baby carrot, the crunchiness of a cucumber slice, or the blue cheese dressing. Maybe it's the way the light of a particular bandwidth of frequencies reflects from a uniquely wrinkled green pea buried under a crisp romaine leaf.

"Quantum physics" didn't make possible toast, fireworks, or the weird light from bubbles any more than your Uncle Buck did.

Tim LaFave Jr
(tlafave@buffalo.edu)
University at Buffalo
Buffalo, New York

Correction

April 2023, page 32—The acknowledgment should read, "Work on 'Multimodal studies to understand pregnancy and prevent stillbirth' is supported by Wellcome Leap as part of the In Utero Program." PT

CONTACT PHYSICS TODAY

Letters and commentary are encouraged and should be sent by email to **ptletters@aip.org** (using your surname as the Subject line), or by standard mail to Letters, PHYSICS TODAY, American Center for Physics, One Physics Ellipse, College Park, MD 20740-3842. Please include your name, work affiliation, mailing address, email address, and daytime phone number on your letter and attachments. You can also contact us online at **https://contact.physicstoday.org**. We reserve the right to edit submissions.

How a cloud of insects is (and isn't) like a magnet

The renormalization group, a powerful method that uses the tools of quantum field theory, has found a place in biophysics.

You don't have to know the position and momentum of every molecule in the North Atlantic Ocean to study the Gulf Stream current, but small- and large-scale phenomena are still connected. That's especially true in the vicinity of a phase transition. If you want to understand what's going on in water near its critical point, where the distinction between liquid and gas fades away, or in a magnet near its Curie temperature, at which permanent magnetization vanishes, you can't confine your investigation to just one length scale. They're all important.

Nowadays it may be tempting to approach such a problem by brute force: Set up a computer simulation sufficiently intricate to capture a wide range of scales, and let it run. But half a century ago, when that was a less realistic option, physicists were driven by necessity to seek a more analytically tractable solution. The result, the renormalization group (RG), is the art of mathematically zooming out—of blurring over a system's finest details and compressing everything else to a smaller scale—in search of scaling laws and self-similar phenomena. When applied across diverse areas of physics, spanning both condensed matter and high energy, RG calculations have identified deep connections among ostensibly dissimilar systems.[1]

The RG's reach now includes biology, thanks to new work by researchers led by Andrea Cavagna, of the Institute for Complex Systems of the National Research Council in Rome, who applied the method to a dynamical field theory of insect swarms.[2] Their theoretical results agree extremely well with both numerical simulations and experimental observations. It's not the first attempt to perform an RG calculation on a living active system. But the successful comparison with experiment is new.

Cloud computing

Cavagna and colleagues have long been interested in collective biological behaviors. They were initially inspired by the flocking starlings that put on especially impressive displays in their home city of Rome (see PHYSICS TODAY, October 2007, page 28). A flock of thousands of birds can undulate and swirl in near unison, despite having no one leader. The collective synchrony emerges from each bird's tendency to fly in the same direction as its neighbors.

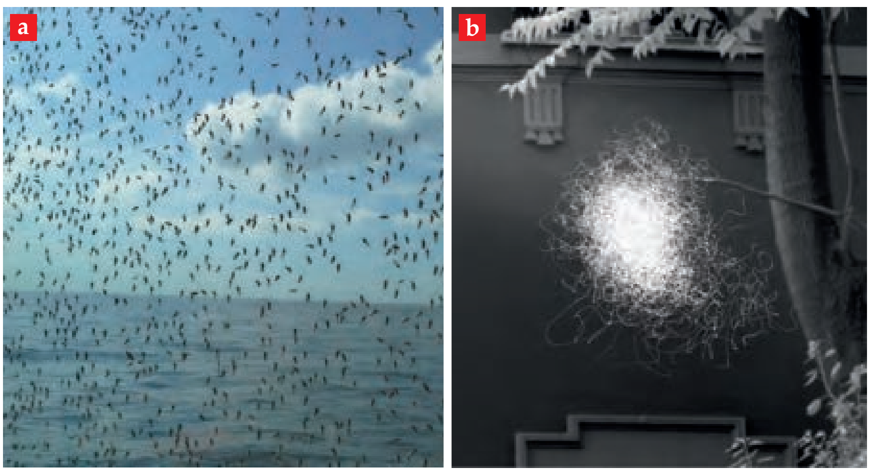


FIGURE 1. SWARMING INSECTS buzz around in all directions while still following directional cues from their neighbors. The result is a pattern of correlations that lends itself to treatment by some of the most sophisticated tools of statistical physics. (a) A swarm of midges photographed from a research vessel on Lake Erie. (b) A composite trajectory of some 300 midges recorded in a park in Rome. (Panel a courtesy of NOAA; panel b adapted from ref. 4.)

Swarming insects, in contrast, don't all fly in the same direction: The cartoon depiction of a horde of angry wasps determinedly chasing down their fleeing victim is a myth. Rather, a real swarm tends to hover in place like a cloud, with its constituent insects buzzing to and fro, as shown in the photo in figure 1a and the composite trajectory image in figure 1b.

Nevertheless, swarming insects do imitate their neighbors—just not strongly enough to create any swarm-wide order. In physics parlance, swarms and flocks are the disordered and ordered phases of the same system, much like a magnet above and below its Curie temperature.

Whatever the temperature, a magnet's spins are governed by the competition between an energetic preference to align with their neighbors and an entropic preference to orient randomly. Below the Curie temperature, energy wins, and the spins align. Above the Curie temperature, entropy wins, and there's no net magnetization.

But something special happens in a magnet that's only barely above its Curie temperature. Clumps of aligned spins emerge that, although smaller than the whole magnet, can be quite large. In fact, they span all length scales—no matter how large the magnet, there's a good chance that it contains a clump that's a significant fraction of its size—the very problem that necessitated the invention of the RG to explain systems near phase transitions.

From their initial observations of insect swarms in the wild, Cavagna and colleagues found a marked resemblance to barely demagnetized magnets.[3] The swarms lacked a single preferred direction, but plenty of clumps of insects momentarily aligned their motion. The larger the swarm, the larger the clumps—the hallmark of a scale-free, near-critical system.

A magnet's Hamiltonian is easy to understand, even if its near-critical be-

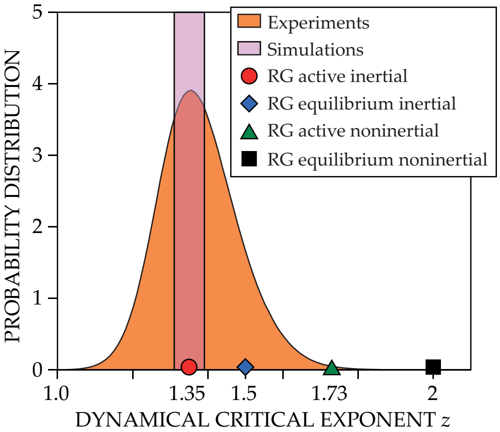


FIGURE 2. DYNAMIC SCALING, the relationship between a system's correlations in space and in time, is quantified by the critical exponent *z*. For observed and simulated insect swarms, *z* closely matches the value predicted by the renormalization group (RG) applied to a theory that incorporates activity and inertia. The RG calculations on other theories yield values of *z* significantly larger. (Adapted from ref. 2.)

havior is not. Could an insect swarm—a collection of living animals, each of which can sense and respond to all the complexities of its environment—be governed by a similarly simple set of rules? "The space and time correlation functions were horrible beasts," says Cavagna, "but in physics, you can make a big simplification and describe everything with one exponent." The dynamical critical exponent *z* quantifies how correlations in space scale with correlations in time. In 2017 the researchers tried applying the dynamical scaling hypothesis to their swarm data. It worked.[4]

Order of life

But the value of *z* for the swarms was extremely low: 1.37 ± 0.11, as shown by the orange bell curve in figure 2. For magnets, in contrast, *z* is exactly 2, shown by the black square. Intuitively, *z* can be thought of as a measure of how fast fluctuations spread across a system, with smaller values representing faster spread and a fundamental speed limit at $z = 1$. Insect swarms were not only faster than magnets at propagating fluctuations, they were also faster than any existing theory could explain.

Two key ingredients, the researchers found, distinguish a swarm from a magnet and cause fluctuations to propagate faster in the former. The first is activity: Whereas a spin in a magnet always interacts with the same set of neighbors, insects move under their own power, so their nearest neighbors are constantly changing. The second is inertia: Insects don't react immediately to what their neighbors are doing, so small-scale heterogeneities in a swarm can persist for some time. In other words, if a swarm is viewed as a fluid, it's one without a lot of viscous drag.

Activity and inertia had both been treated separately in RG calculations before. The equilibrium (nonactive) inertial model was one of the classic RG successes[5] from the 1970s: It describes the behavior of superfluids, among other things, and has a *z* of 1.5, shown by the blue diamond in figure 2. The active noninertial model didn't emerge until later, once interest in the physics of active matter—that is, living things—had begun to gain traction as a field of study.[6] It gives $z = 1.73$, shown by the green triangle.

Encouraged by how activity and inertia each push *z* in the right direction, the researchers began to think about applying the RG to a model that incorporates both. When they embarked on the project in the summer of 2019, it was a two-person effort of Cavagna and PhD student Luca Di Carlo. But the complexity quickly spiraled out of control.

In simple terms, an RG calculation entails renormalizing an equation in two steps—coarse-graining and rescaling—and hoping to get a new equation of the same mathematical form. When that happens, the calculation is finished: The equation describes how the system behaves self-similarly across all length scales.

Often, however, renormalizing an equation gives a more complicated equation. "When that happens, the RG is telling you, 'Hey dummy, you forgot to put a term in your equation,'" says Cavagna. Renormalizing the more complicated equation may complete the process, or it may yield an even more complicated equation. And so on.

The active inertial model has more ways than usual for its equations to become complicated. Roughly speaking, in the viscous regime, the social force—how individuals influence their neighbors—is proportional to the first time derivative of velocity. But in the inertial regime, it's proportional to the second time derivative. The additional derivative introduces more ways to make nonlinear terms, which complicate the equations.

The RG calculation was soon yielding more terms than could feasibly be calculated by hand. "It was a nightmare," says Cavagna. Only after Di Carlo and Mattia Scandolo, another student in the group, developed a specialized Mathematica code to help with the calculation did they start to get a handle on the renormalization. "I was very close to giving up," says Cavagna, "but they made the final push to get it done. I could never have done it without them."

The result: $z = 1.35$, shown by the red circle in figure 2, and in excellent agreement with the data. The researchers wrote up their work that summer and submitted the paper for publication. When the reviews came in, says Cavagna, "they said, 'This looks great, but it could just be an accident. You need to do numerical simulations for us to believe this is real.'" That task fell to Giulia Pisegna, another student who'd worked on the project. Even though she'd already left the group for a postdoctoral position, she set up the large, intricate simulations, let them run, and found $z = 1.35 \pm 0.04$, shown by the pink stripe.

Even with the convergence of theory, simulation, and experiment, *z* is still just one quantity. The active inertial model successfully predicts other properties of insect swarms, says Cavagna, "but we'd really be happier if we had more exponents." More critical exponents exist than just *z*, and the RG is capable of calculating them. The limitation now is in the experimental data. Some critical exponents, for example, manifest themselves only in a system's response to an external stimulus, such as a magnet in an applied magnetic field, and it's not yet clear what the equivalent experiment would even be on an insect swarm. "Until we have data, I'm not sure we want to invest all the time to do the calculations," says Cavagna. "As we say in our group, 'No data, no party.'"

Not dead yet

The applications to biology have the potential to breathe new life into the RG. A major theme in the story of RG calculations has been about finding the same critical exponents and behaviors across disparate physical systems. Nevertheless, the $z = 1.35$ regime is completely new. (So, for that matter, was the $z = 1.73$ regime of the active noninertial model.) Finding new critical phenomena in active systems may be a way that new physics lurks in living matter (see the article by Paul Davies, PHYSICS TODAY, August 2020, page 34).

"But fewer and fewer people remember how to do these calculations," says Cavagna. The sophisticated math of the RG builds on the methods of quantum field theory, which is increasingly seen as a merely optional part of the physics curriculum, especially for would-be biophysicists. "Biophysics is so rich in amazing and fascinating problems," says Cavagna. "But we have so many students coming to us wanting to study the birds, and we have to tell them that they can't because they don't know field theory. And you really have to learn it when you're young—trying to pick it up when you're 35 or 40 is too late."

Johanna Miller

References

1. K. G. Wilson, *Sci. Am.* **241**(2), 158 (1979).
2. A. Cavagna et al., *Nat. Phys.* (2023), doi:10.1038/s41567-023-02028-0.
3. A. Attanasi et al., *Phys. Rev. Lett.* **113**, 238102 (2014).
4. A. Cavagna et al., *Nat. Phys.* **13**, 914 (2017).
5. P. C. Hohenberg, B. I. Halperin, *Rev. Mod. Phys.* **49**, 435 (1977).
6. L. Chen, J. Toner, C. F. Lee, *New J. Phys.* **17**, 042002 (2015).

Macroscopic mechanical oscillator is herded into a Schrödinger cat state

A new demonstration of a famous thought experiment has coaxed a classical 16 μg cluster of atoms to behave like a quantum object.

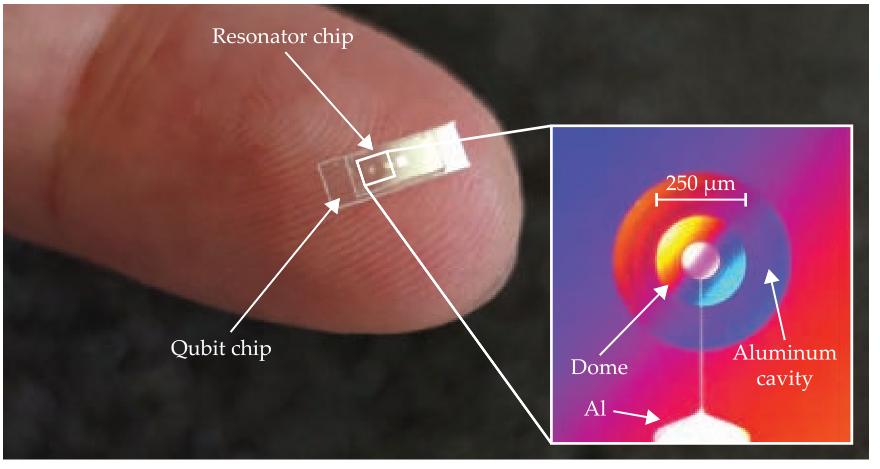


FIGURE 1. A HYBRID QUANTUM DEVICE is made of a superconducting qubit coupled with a classical mechanical resonator, each fabricated onto a sapphire crystal chip. The optical microscopy image (inset) shows a dome of piezoelectric aluminum nitride that strongly couples the two parts of the device, which are placed within a superconducting aluminum cavity. The device generated a Schrödinger cat state—a quantum superposition of, in this case, 10^{17} atoms in the resonator wiggling in two opposite directions simultaneously. (Courtesy of Matteo Fadel.)

Cats and other everyday things we see don't appear to exist in two mutually exclusive states at the same time. But Erwin Schrödinger's thought experiment in 1935—that a cat could be dead and alive simultaneously—and subsequent quantum theory specify that a superposition of two states should be observable even for macroscopic objects. But despite its superlative predictions of phenomena at the smallest scale of matter, large objects don't appear to exhibit quantum behavior. Does that mean the theory breaks down at some microscopic–macroscopic transition, beyond which quantum mechanics ceases to apply? (For more on the classical–quantum boundary, see PHYSICS TODAY, May 2004, page 25.)

One way to test whether quantum mechanics is valid at macroscopic scales is to experimentally generate a cat state, named after Schrödinger's thought experiment. It's a quantum state that is a superposition of two classically distinct states at the same time. In phase space, the two states correspond to well-separated probability distributions.

When it comes to determining what counts as macroscopic, researchers don't entirely agree on a formal definition. But there are two generally accepted criteria that most would say qualify a cat state as macroscopic: The system being studied should be large—often, although not exclusively, measured by its number of atoms—and the two superposed states should have a distinguishable difference, such as being dead and alive or being separated by a long distance.

Various explanations have been put forth for why macroscopic cat states have never been observed. Perhaps macroscopic objects interact with their environments in such complex ways that no quantum state can survive for any measurable coherence time (see the article by Wojciech Zurek, PHYSICS TODAY, October 2014, page 44). Or such objects may have intrinsic sources of noise that interfere with the generation of quantum states.

Nevertheless, numerous experiments have demonstrated Schrödinger cat states at sizes that are approaching macroscopic scales, in trapped-ion quantum computers, superconducting quantum interference devices, Bose–Einstein condensates, and matter–wave interferometers. Now Marius Bild, Matteo Fadel, Yu Yang, and colleagues—all part of ETH Zürich's Hybrid Quantum Systems Group, led by Yiwen Chu—have created a cat state in a mechanical resonator made of 10^{17} atoms, which is the most massive demonstration to date.[1]

Classical–quantum hybrid

The challenges of making cat states are similar to the difficulties that any researcher encounters when studying a system's quantum behavior. The choice of how much to isolate a quantum system from the environment must be balanced against the efforts to measure it. If a quantum system were completely isolated from its surroundings, no environmental noise would disturb it, and its state would never be measurable. Although cat states and other quantum systems can be well

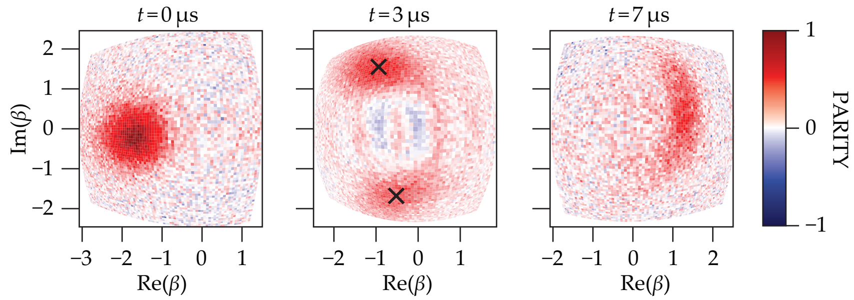


FIGURE 2. WIGNER-FUNCTION MEASUREMENTS show the quantum behavior in a phonon state created in a mechanical resonator coupled to a superconducting qubit. The real and imaginary parts of the complex displacement amplitude β are measured in parity units. The left and right panels show the quantum system immediately before and after the formation of the Schrödinger cat state, shown in the middle panel. It's characterized as a superposition of oscillations from the resonator that have displacements in opposite directions. The two components of the cat state (black crosses) have interference fringes in between, which indicates that the two are in a coherent quantum superposition. (Adapted from ref. 1.)

isolated, the size of macroscopic ones makes them particularly sensitive to noise.

To balance those issues, Chu and her colleagues have been working over the past several years on developing hybrid quantum systems that couple a classical solid-state material to a quantum device.[2] That approach allows not only for the system to be relatively large but also for the quantum half of it to be well controlled and isolated. Rather than focus on inducing quantum behavior in light, atoms, or other commonly studied platforms, the group at ETH Zürich has been designing an acoustodynamical system.

One part of the device is a superconducting transmon qubit. It's a type of two-state quantum system with a well-defined charge—in this case, from Cooper pairs of superconducting electrons.[3] Transmons, and superconducting qubits in general, are less sensitive to noise than some other qubits and are relatively easy to control and measure (see Physics Today, July 2009, page 14).

The classical half of the hybrid quantum system is an acoustic-wave resonator fabricated on a sapphire crystal. Coupling the resonator to the superconducting qubit is a dome of piezoelectric aluminum nitride, which converts the qubit's electrical signal into quantum mechanical oscillations of the resonator. The dome needs to strongly couple the qubit to the mechanical mode, but at the same time, it must also confine the mechanical oscillations in a well-defined volume so they can be measured.

The cat state is created when sound waves generated by the piezoelectric material interact with the qubit and induce a superposition of two opposite-phase oscillations of atoms in the crystal lattice. Those oscillations are phonon modes—collective excitations of all the atoms that can be thought of as quantized sound waves. The entire apparatus, shown on a finger in the photo in figure 1, sits in an aluminum cavity that isolates the qubit from environmental noise.

Cat out of the bag

In classical mechanics, a particle can be represented by a point in phase space with a definite position and momentum. But because of the Heisenberg uncertainty principle, that approach doesn't work for quantum phenomena. Instead, a quantum state can be represented by a quasi-probability distribution in phase space, called a Wigner function, that gives a description of the system's state.

To confirm the existence of the cat state, Chu and her colleagues measured the Wigner function of the phonon state. Unlike classical probability distributions, it can take on positive and negative values. Because negative values of the Wigner function have no classical interpretation, seeing them means that the system has an intrinsic quantum mechanical nature rather than some sort of statistical mixture, such as having a cat be dead 50% of the time and alive for the other 50%.

The researchers first allowed the resonator and qubit to interact and become entangled, and the left panel of figure 2 shows the measured Wigner function for the initial phonon state. After about 3 μs, the cat state emerged in the coupled qubit–resonator system. As shown in the middle panel, the cat state is characterized by two distinct components with interference fringes between them—the telltale sign that the two components are in a coherent quantum superposition.

The excitation in the mechanical resonator lasted for as long as 80 μs and consisted of atoms in the resonator's crystal lattice wiggling in two directions with opposite displacement amplitudes of about 2.1×10^{-18} m, which is a fraction of a proton's radius.

Although that subnuclear distance is much shorter than in other demonstrations of cat states,[4] Chu and her colleagues are just getting started, and they expect to see mechanical cat states with larger amplitudes in the future. Markus Arndt of the University of Vienna says that "the system of circuit quantum acoustodynamics is pretty unconventional. I had not expected to see a Schrödinger cat realized in such a system, and I admire what they have achieved."

Sound bytes

Seeing if it's possible to generate macroscopic cat states is good for more than testing how quantum mechanics works at large scales. Such states are also useful for various quantum technologies, including error-protected processing of quantum information.

Mechanical resonators have greater storage capacity compared with electromagnetic resonators and other quantum-information storage devices. Because the speed of sound is slower than the speed of light, a mechanical resonator has more motional degrees of freedom that are available for storing quantum information, per unit volume. That advantage, however, is limited for now at least because of the coherence time. In the mechanical system by Chu and her colleagues, the coherence time is about an order of magnitude less than competing microwave-cavity technology.

But Chu and her colleagues think they're capable of closing that gap. "I definitely don't think we've hit the upper limit," says Chu. "People have made better qubits and better resonators. I think for us, the challenge is to maintain those properties, make improvements, and then maintain them once we put everything together."

Alex Lopatka

References

1. M. Bild et al., *Science* **380**, 274 (2023).
2. Y. Chu et al., *Science* **358**, 199 (2017).
3. J. Koch et al., *Phys. Rev. A* **76**, 042319 (2007).
4. Y. Y. Fein et al., *Nat. Phys.* **15**, 1242 (2019). PT

Free textbooks and other open educational resources gain popularity

Affordable and customizable, they contribute to making higher education more inclusive and accessible.

Given a choice between paying $200 for a textbook or taking an equivalent course with a free textbook, what would you do?

The prices of college textbooks have skyrocketed: From 2011 to 2018, they went up by 40.6% in the US, according to the Bureau of Labor Statistics' Consumer Price Index. That can add up to as much as $1000 for a single semester. So it's no surprise that freely available, openly licensed textbooks, lectures, simulations, problem sets, and more—known collectively as open educational resources (OERs)—are having a moment.

Last year, for the first time, more than half of US college faculty reported "some level of awareness" of OERs, finds Bay View Analytics, a company that conducts research at the intersection of technology and education.

A major player on the OER scene is OpenStax. Based at Rice University, the initiative currently has more than 60 titles focused mostly on large introductory college courses. "They have revolutionized the OER market," says Jeff Seaman, Bay View's director. OpenStax put out its first textbook in 2012, for algebra-based physics. In spring 2022 OpenStax announced that its introductory astronomy textbook was the US leader. In less than a decade, several OpenStax textbooks have risen to third through fifth place in their markets, says Seaman. "No other publisher has had that growth." From its origins in introductory-level college classes, OpenStax is moving down to the high school and up to upper-division levels.

Around the turn of the millennium, various forms of OERs got started, including Wikipedia; shared lectures, such as MIT's OpenCourseWare; and online science simulations from the University of Colorado Boulder known as PhET. OpenStax grew out of a predecessor that Richard Baraniuk, a professor of electrical and computer engineering at Rice, started in 1999. At the time, he wanted to offer his students a better, customizable textbook. "Instead of thinking of books as glued-together pieces of paper," he says, "we could think of them as Lego bricks" by taking advantage of the Web.

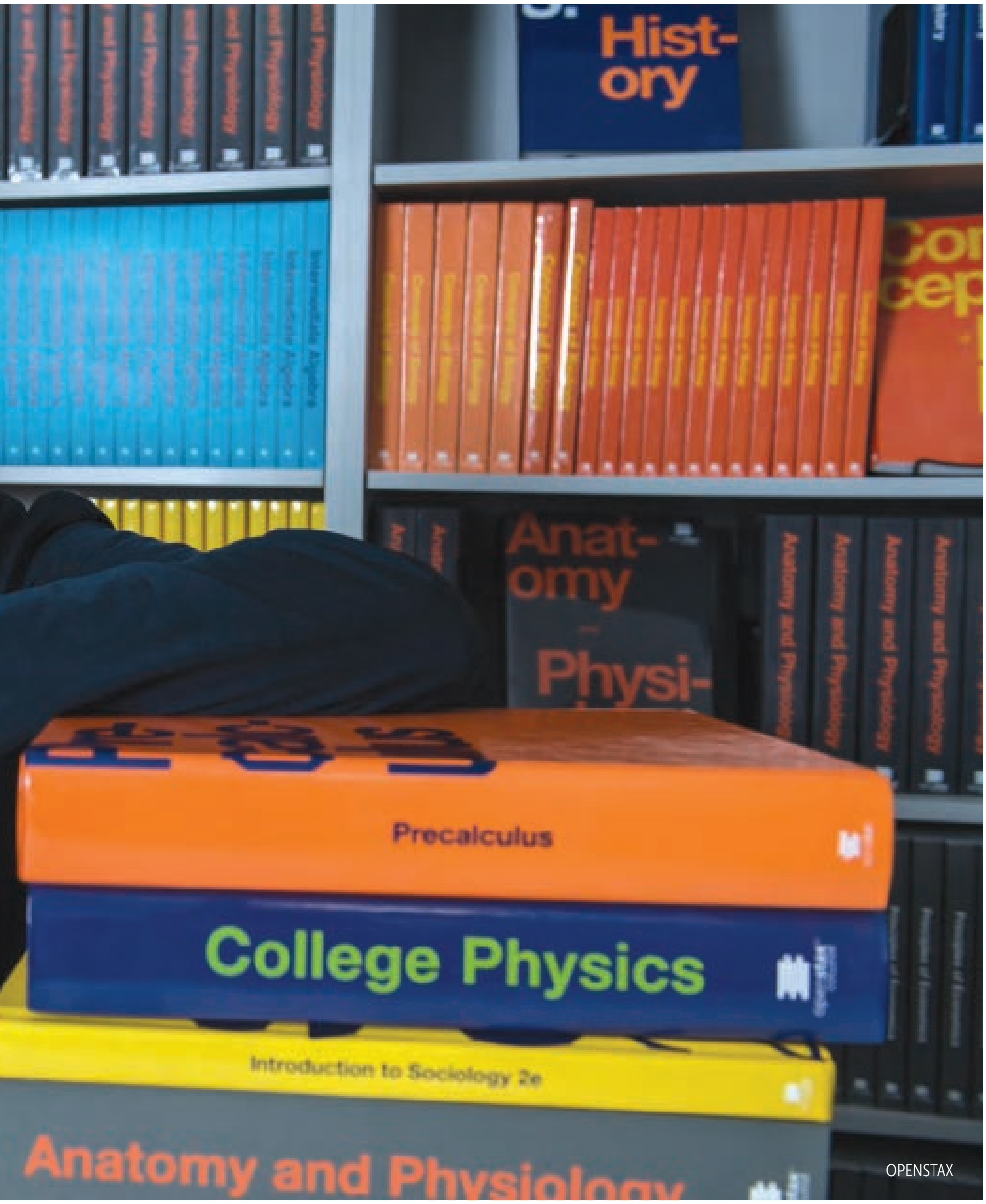
Free, open textbooks have "taken off much faster than I expected," says Baraniuk. He estimates that to date, students have saved nearly $2 billion by using OpenStax books instead of buying comparable traditional textbooks. And being free is not the only selling point for open textbooks and OERs more broadly: In the past few years, says Lauren Woolsey, who teaches physics and astronomy at Grand Rapids Community College in Michigan, "the messaging has shifted to the role they can play for social justice and equity."

Evolution revolution

"Our resources are extensively peer reviewed and copyedited, and interactive elements are checked," says OpenStax editor-in-chief David Harris, who previously worked at three major traditional publishers. OERs are typically published under a Creative Commons license, which can take several forms but generally allows material to be legally used, adapted, translated, and shared by anyone anywhere. Because it's open source, says Harris, "we have to create most of the lavish drawings and illustrations ourselves." The total cost to make a new textbook comes to $500 000 to $1 million, he says. OpenStax is funded mostly by philanthropies, with the William and

Flora Hewlett and Bill and Melinda Gates Foundations being among the main givers.

OpenStax books are similar in content and scope to their traditional counterparts, Harris says. One reason is to ease the switch to OERs for instructors and institutions. But there are differences: As digital textbooks, OpenStax publications can be updated regularly and cheaply. And unlike digital textbooks from traditional publishers, users can download searchable electronic versions, which is especially helpful if their internet is unreliable. (About 7% of students order print versions of OpenStax textbooks, the publisher says. They cost $25 to $50—much less than tradi-



OPENSTAX

OPENSTAX FOUNDER AND DIRECTOR RICHARD BARANIUK with a selection of hard-copy versions of open textbooks. The publisher says that more than 23 million students have used its resources since 2012.

tional textbooks—to cover printing expenses.) And OpenStax and other open textbooks often link to open-source educational materials, including videos and such interactive simulations as PhET. Many of the books have an associated online site where instructors can share ancillary materials.

OpenStax authors are paid a fixed amount—there are no royalties. Few authors strike it rich writing textbooks, says Harris. "Under our model, everyone understands how much they'll be paid." The vast majority do it for the OER mission, he adds. "We have authors coming to us."

Andrew Fraknoi, a professor at the Fromm Institute of the University of San Francisco, is lead author of the OpenStax astronomy textbook. "Teaching introductory astronomy is one of the most profound ways that the astronomy community influences the public," he says. "Students will be the next generation of teachers, congressional representatives, taxpayers, and parents. This is important work."

The digital versions of traditional textbooks—the top publishers are Pearson, Cengage, and McGraw-Hill—are cheaper than their hard-copy counterparts. And commercial publishers are trying to stem their loss of market share to open textbooks. For example, most of them offer "inclusive access" packages. They offer multiple textbooks and bundle electronic textbook access together with quiz and homework platforms. Such packages are often wrapped into tuition fees, making textbook costs difficult to disentangle or opt out of. On request, some traditional publishers sell portions of hard-copy books. And they let users rent electronic textbooks individually or as a suite of titles, with leasing times that vary from a single semester to life.

The COVID-19 pandemic helped with the switch to digital textbooks, says Jonathan Perry, a physics instructor at the University of Texas (UT) at Austin. "Why use OER? All textbooks are imperfect. They all have flaws. So why not use the one that is free and accessible?"

Customizing and sharing

Jennifer Kirkey is a physics instructor at Douglas College in British Columbia, Canada. "I'd been sort of aware of OER," she says, but a "light-bulb moment" came about 10 years ago when she realized she could customize the open textbooks. "I was tired of the textbook publishers forcing my pedagogy by what they publish," she says.

Members of Kirkey's department jointly created their own versions of OpenStax textbooks. They slashed a lot, turning 30 chapters into 10, she says. "Students can be scared, math phobic, so being able to give them a smaller book was better than saying, 'Skip these sections.' "

They also significantly broadened the cultural context in both the astronomy and algebra-based physics OpenStax textbooks, Kirkey says. "We added examples from many cultures, including First Nations of Canada, to make them more visible. We removed the implicit assumptions that everyone knows minutiae about baseball and football." The US versions talk about American astronauts, she adds. "We include examples from Canada and other countries."

Grand Rapids Community College's Woolsey says that using OERs improves her teaching. "I don't feel tethered. I am

STUDENTS AT DOUGLAS COLLEGE in British Columbia regularly hold "tabling" events to advocate for free textbooks. They collect students' textbook receipts and shout out to professors who use open textbooks.

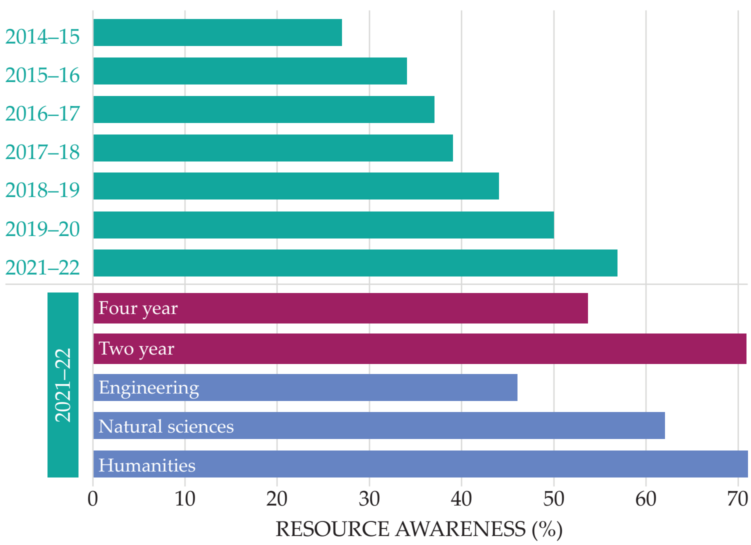
more free with assignments and in how we veer into material that comes up," she says. "If I were assigning a $200 textbook, I'd feel I had to use it enough to justify the cost."

Woolsey posts slides, videos, and problem sets on the astronomy instructors' hub at OER Commons, a public digital library for educators. "No one has to reinvent the wheel when we have a collaborative spirit," she says. "I like helping other instructors get started. And they have the built-in right to edit."

Affordability, equity, access

Rajiv Jhangiani, a psychology professor at Brock University in Ontario, Canada, has authored open textbooks, advocated for them at the United Nations, and researches their efficacy. When students can see in the course catalog that classes will have a free textbook, enrollment and course completion rates both increase, he says. "Affordability is important for equity and access in higher education."

As long as students have access to the course textbook, their performance doesn't depend on whether textbooks are free, Jhangiani says. But when buying textbooks is a stretch, students go without or resort to pirating, sharing, or pilfering library copies, says Ann Fiddler, director of open education at the City University of New York (CUNY), where she oversees the 25-campus system's OER program. "They come in to the library with razor blades and steal pages from lab books," she says, adding that for some



AWARENESS OF OPEN EDUCATIONAL RESOURCES is growing among US-based higher-education instructors and exceeded 50% in spring 2022. Awareness was highest in the humanities (73%) and lowest in engineering (46%). In the natural sciences, it was 62%. It was also notably higher in two-year institutions (71%) than four-year ones (54%). Data are from annual surveys conducted by Bay View Analytics and averaging 3000 instructors.

Selected resources

- The Open Education Network, https://open.umn.edu/oen, works to make higher education more accessible. It hosts the Open Textbook Library, https://open.umn.edu/opentextbooks, which currently offers more than 1200 open textbooks.
- In addition to providing open licensing, Creative Commons runs programs to train and spread the word about sharing knowledge, https://creativecommons.org.
- "Twenty years of Open Educational Resources: Building robust networks for innovation," by Angela DeBarger and Cathy Casserly for the William and Flora Hewlett Foundation, 7 October 2021, https://hewlett.org/twenty-years-of-open-educational-resources-building-robust-networks-for-innovation.
- The Community College Consortium for Open Educational Resources provides materials and support, https://www.cccoer.org.
- The Mason OER Metafinder conducts real-time searches across many sources, https://oer.deepwebaccess.com.
- The multi-institutional LibreTexts project produces open textbooks, including in physics, https://phys.libretexts.org.
- MIT offers lectures and course materials for thousands of its undergraduate and graduate classes at OpenCourseWare, https://ocw.mit.edu.
- PhET offers a large selection of free, interactive math and science simulations, https://phet.colorado.edu.
- Walter Lewin's lectures on YouTube claim they'll make viewers ❤ physics, https://www.youtube.com/@lecturesbywalterlewin.they9259.
- HyperPhysics explores many physics concepts, http://hyperphysics.phy-astr.gsu.edu.
- Crash Course posts freely available educational videos for high school and college levels on YouTube, https://thecrashcourse.com.
- The oPhysics site offers a collection of interactive physics simulations, https://ophysics.com.
- Some 250 libraries and academic organizations in North America belong to SPARC, a nonprofit advocacy organization that supports open educational resources. It tracks US state policies at https://sparcopen.org/our-work/state-policy-tracking.

students it comes down to choosing between paying for a textbook, food, or rent.

When free textbooks are available for full degree programs, it's known as ZTC, or zero textbook cost. The institution benefits, says Jhangiani. "There are fewer withdrawals and the institution doesn't have to refund tuition."

Although adoption of OERs at the college level is largely up to individual instructors and departments, some systems of higher education and states promote it. The majority of campus OER initiatives are led by libraries. (At the K–12 levels, adoption of OERs is a matter of appealing to schools or school districts.)

British Columbia started an open-textbook project across public higher education in 2012. In 2021 California awarded its community college system $115 million over five years to develop ZTC programs. And in 2016 CUNY was among 38 community colleges nationwide that received grants totaling $8 million from the nonprofit organization Achieving the Dream for the same goal.

New York State gives CUNY and the State University of New York each $4 million annually for OERs. The money goes largely toward incentivizing faculty to convert their courses to OERs, says Fiddler. For a small stipend, many faculty will switch, she says. "And when we come across resistance, we move on. We don't proselytize." So far, 15–17% of courses in the CUNY system use open textbooks, she says. The goal is 100%, she adds. "We've saved more than $100 million in student textbooks in the last five years."

Upping the uptake

Uptake of OERs is growing: In 2021–22, 22% of US faculty surveyed by Bay View said they used OER materials for their largest enrollment course, up from 5% in 2015–16. In other countries, adoption may be higher. MIT's OpenCourseWare and Colorado's PhET, for example, are both used more internationally than in the US.

A barrier to adopting OER materials is inertia—instructors have textbooks and other resources they are used to. Another is the scattered distribution of OERs, which can complicate identifying and finding materials. Others include a mismatch of learning management systems, the narrowed price gap when traditional textbooks include homework management systems, a perception that something that is free won't be as good as something one pays for, and time.

"We might have faculty who quickly locate an OpenStax textbook," says James Glapa-Grossklag, dean of learning resources at College of the Canyons in Santa Clarita, California. "But some of the book doesn't speak to the local curriculum, local needs, or local identities." Adapting a textbook takes time, he continues. "Faculty don't have time, and their institutions usually don't recognize adapting a book as contributing to their tenure packet." That's starting to change, with some institutions, such as the University of British Columbia, revising their tenure and promotion policies to recognize such efforts.

Some instructors and students complain that OpenStax books can be verbose and that they don't offer enough practice problems. Those issues come up with commercial textbooks, too. With OpenStax, though, they could be addressed through adapting the text, the publisher notes, but many instructors don't get around to doing that. "I have not dug in to modify a text," says UT Austin's Perry. "I'd like to, but it's work and time."

Another common concern relates to the paucity of such ancillary materials as question banks and tutorials. That was more of a problem earlier on, says Jhangiani. The OER content producers are giving "more attention in the past few years to ancillary materials."

The goal of OERs is to provide greater access to higher education for more people, says Glapa-Grossklag. "That goal contributes to a functioning democracy and equitable society."

Toni Feder

Capture alone isn't sufficient to bottle up carbon dioxide

The US has practically boundless capacity to store carbon dioxide. It just needs to find a way to do it.

Billions of dollars in federal subsidies, combined with enhanced tax credits, have stimulated dozens of new carbon capture and storage (CCS) and direct air capture (DAC) projects in the US. But a severe bottleneck in the regulatory process for transport and storage is slowing progress toward meeting the goal of the Biden administration to reach net-zero emissions by 2050.

The 2021 bipartisan Infrastructure Investment and Jobs Act (IIJA) appropriated more than $10 billion in R&D and demonstration project funding over five years for CCS and DAC. (See Physics Today, January 2022, page 22.) And last year's Inflation Reduction Act provided a powerful incentive to projects that lock up carbon dioxide: raising the tax credit for capture and permanent storage of CO_2 to $85 per ton from its previous level of $50. The credit was elevated from $35 to $50 per ton if the CO_2 is utilized for oil production or other applications, such as incorporating it into concrete and building materials.

In May, the Environmental Protection Agency published a notice of proposed rulemaking that, if it becomes final, would establish CO_2 emission limits on existing and new fossil-fuel power plants. That could stimulate further demand for CCS in the years ahead.

An April report from the Department of Energy says models suggest that achieving the US net-zero emissions goal will require capturing and storing 400–1800 million tons of CO_2 per year by 2050. Today the US leads the world in capturing around 20 million tons annually, almost all of it from ethanol refining, hydrogen production, and natural-gas processing. Those sources provide concentrated, high-purity CO_2 streams, costing as little as $20 per ton to capture, according to the International Energy Agency. Most of that gas is transported in supercritical form via pipelines to depleted oil reservoirs for use in enhanced oil recovery. Although the injected CO_2 will stay locked up, the petroleum that it pushes out will eventually add to the atmospheric CO_2 burden.

THE PETRA NOVA plant in Thompsons, Texas, was the only US power plant to successfully capture carbon dioxide from postcombustion at a commercial scale. The CO_2 was compressed and piped to an oil field for enhanced oil recovery. The system operated from 2017 until 2020, when low oil prices made capture uneconomic.

Carbon capture from industries with lower-purity CO_2 streams and distributed process emissions—power, ammonia, cement, steelmaking, and other plants—will require improved project economics, the DOE report says. Lowering costs in those industries, which the International Energy Agency says range up to $120 per ton, could be accomplished through a combination of technology improvements, tighter regulations, and enhanced subsidies. In any case, widespread decarbonization of those industries isn't likely to come before 2030, the DOE report says.

Carbon-capture projects in the US totaling around 140 million tons per year are expected to be in place by 2030, according to BloombergNEF. Yet as capture technologies are scaled up and deployed to collect millions of tons of CO_2 annually, their success will be contingent on the more mundane tasks of building out today's tiny CO_2 storage and transport infrastructure. Public concerns about pipeline safety and possible leakage from underground storage have the potential to slow those deployments.

Credits and subsidies

On 17 May DOE announced awards totaling $251 million to 12 projects in seven states to expand the CO_2 transportation and storage infrastructure. It's the first tranche of the $2.5 billion that was appropriated over five years by the IIJA specif-

ically for those purposes. Nine projects totaling $242 million will support new and expanded commercial carbon storage projects with capacities to store 50 million tons of CO_2 or more. Three projects for a total of $9 million were selected to prepare detailed engineering design studies of proposed regional CO_2 pipeline networks to carry captured gas from power plants, ethanol refineries, and other industrial operations either to permanent storage or to factories where it will be incorporated into long-lived products such as concrete.

In its May announcement, the department also solicited proposals for additional storage and transportation projects to be funded from the remaining $2.2 billion in dedicated IIJA funding.

Because capturing CO_2 from ambient air is far more difficult than scrubbing it from concentrated point sources, larger tax credits are provided for DAC: $180 per ton if the CO_2 is permanently stored and $130 per ton if the CO_2 is utilized. The few commercial DAC plants in operation have costs well over $200 per ton.

The IIJA provided $3.5 billion to help finance four regional DAC hubs. DOE expects to announce awards totaling $1.2 billion of that amount this summer.

In April the Occidental Petroleum subsidiary 1PointFive broke ground on a DAC plant in West Texas that aims to capture and store 500 000 tons of CO_2 per year beginning in 2025. The company is currently seeking permits for 1.2 billion tons of geological storage capacity along the Gulf Coast. Plans call for building 70 similar plants around the globe by 2035, and more if further policy incentives are enacted.

The largest DAC project announced to date is CarbonCapture's Project Bison, the initial stages of which the company hopes to begin operating late this year. The company says that by 2030 the modular plant will suck 5 million tons of CO_2 per year from the air—the amount emitted by an average coal power plant—to be injected into a saline aquifer at its site in Wyoming.

A choke point

In addition to technology development and financing, the rate at which CCS and DAC projects are deployed will hinge on the EPA's issuance of permits required for CO_2 injection wells. To date, the EPA has been agonizingly slow to approve the so-called Class VI permits, critics say.

The US has sufficient geologic storage capacity for trillions of tons of CO_2, enough to store the entirety of the nation's emissions for hundreds of years, according to DOE. But the porous rock formations and deep saline aquifers that are ideal for permanent storage aren't spread uniformly across the country, and each injection site must be carefully characterized to ensure that it can retain the CO_2 for 10 000 years or more.

Today just two US facilities are capturing and permanently storing appreciable amounts of CO_2 in a geological formation: the Decatur, Illinois, ethanol refinery operated by Archer Daniels Midland and the Red Trail Energy ethanol plant in North Dakota. Archer Daniels Midland injects the captured gas into the only two CO_2 injection wells that are permitted by the EPA. Those Class VI permits are distinct from permit types that are issued for oil and gas drilling, and hazardous and oil and gas drilling waste disposal.

Class VI permit applications are piling up fast at the EPA. In June 2022 there were 14 in the queue. "A few weeks ago, there were 75, and when I checked this morning, there were 87," said Danny Broberg, associate director of the Bipartisan Policy Center, on 25 May. "We need to significantly scale up how we can do CO_2 injection without sacrificing environmental principles."

"The stark reality is that any chance of meeting our net-zero emission goals by 2050 is going to require significant carbon capture build-out in this country," says Jeremy Harrell, chief strategy officer of ClearPath, a Washington, DC, advocacy group. "We need fundamental permit reform to reduce the sheer amount of time it takes to permit a Class VI site."

The EPA's review of the Archer Daniels Midland permit applications took six years. In an October 2022 report to Congress, the EPA said it expects to process the applications within two years of their submission. But the agency cautioned that the exact time frame depends on the quality and quantity of site-specific data submitted, the amount of time the applicant takes to respond to requests for additional information, and the number and complexity of public comments received on the draft permit.

Jessie Stolark, executive director of the advocacy group Carbon Capture Coalition, says the EPA has made some progress in the year since passage of the IIJA, which provided the agency $25 million to bolster the permit review process. "We do know that standing up and staffing a regulatory program of this nature and complexity takes time, including addressing critical staffing needs at the agency," she adds.

Several states have asked the EPA to delegate the Class VI well-permitting process to them to accelerate the approval process. Wyoming and North Dakota were granted such authority, known as primacy, during the Trump administration. Four other states have applied for Class VI primacy, but none have yet been approved by the current administration. The IIJA provided $50 million to help states establish Class VI primacy programs. Not every state is expected to apply; unfavorable storage geology exists in parts of the Southeast, for example, notes Harrell.

The EPA announced in April its intention to delegate Class VI authority to

AN INJECTION WELL in Mississippi was sponsored by the Department of Energy to test the effectiveness of injecting and storing carbon dioxide in a deep saline reservoir.

Louisiana. That announcement came soon after Senator Bill Cassidy (R-LA) threatened to hold up confirmation of Biden's EPA nominees. ClearPath estimates that 6 million tons of CO_2 per year could be locked up if Louisiana's 10 pending Class VI permits are approved. CF Industries, for example, has plans to capture 2 million tons of CO_2 each year from its ammonia plant, transporting it through an existing pipeline for storage beneath an injection site owned by ExxonMobil.

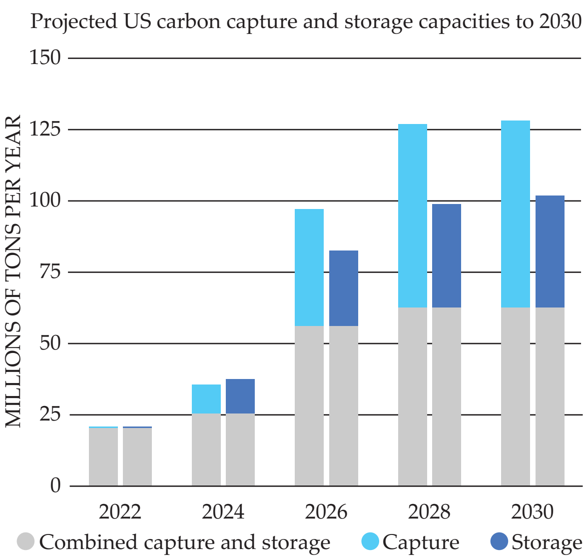
In 2022, Red Trail Energy began injecting about 500 tons of CO_2 per day more than a mile below the surface in North Dakota, becoming the first to operate a Class VI well approved under a primacy agreement. The state approved the permit in nine months. A second Class VI permit, to store CO_2 that is to be captured from a coal power plant owned by the electric cooperative Minnkota, is nearing approval by the state.

North Dakota was able to act quickly on permits because the oil and gas industry had already well characterized the state's geology, says Harrell. That should also apply for Texas, which submitted its primacy application in December 2022. Arizona and West Virginia also have primacy applications pending.

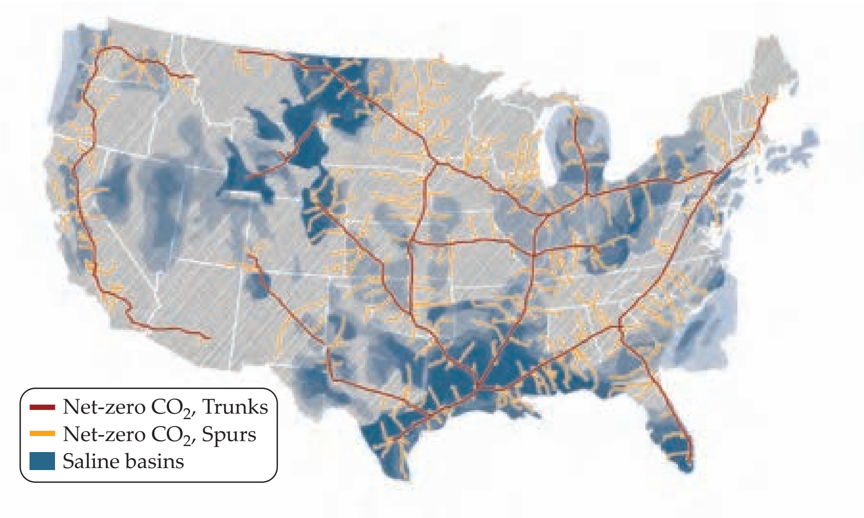
Transport

Virtually all of the 8050 km of pipeline built for CO_2 transportation has been for supplying oil fields that have enhanced oil recovery operations. One scenario modeled by the advocacy group Great Plains Institute shows that moving 300 million tons of CO_2 per year from emissions sources in the Midwest, South, and Great Plains states to deep underground storage sites would require around 46 700 km of pipelines. But many net-zero modeling analyses say the US will require a gigaton of storage by 2050. ClearPath estimates that would require a network of more than 100 000 km.

At least two companies are proposing overlapping pipelines across Midwestern states to collect CO_2 from existing ethanol and fertilizer plants for injection in Illinois



CARBON DIOXIDE CAPTURED in the US will soon outstrip the capacity to store it, according to projections from the International Energy Agency. The plot shows all operating and planned CO_2 capture, transport, storage, and utilization projects commissioned since the 1970s that have an announced capacity of more than 100 000 tons per year, or 1000 tons per year for direct-air-capture facilities. (Based on ClearPath's analysis of the National Carbon Sequestration Database and Princeton University's Net-Zero America project.)



NEW PIPELINES will be required to accommodate the volumes of carbon dioxide needing to be captured and stored if the US is to meet the goal of the Biden administration to attain net-zero emissions by 2050. Shown here is one illustrative network from the nonprofit advocacy group ClearPath.

or North Dakota. Navigator CO_2 Ventures, based in Omaha, Nebraska, proposes to build a 2000 km interstate pipeline network to transport up to 15 million tons annually from more than 30 ethanol and fertilizer manufacturers across five Midwestern states. Summit Carbon Solutions plans 3200 km of pipeline to gather and store CO_2 from more than 30 ethanol plants in the same region.

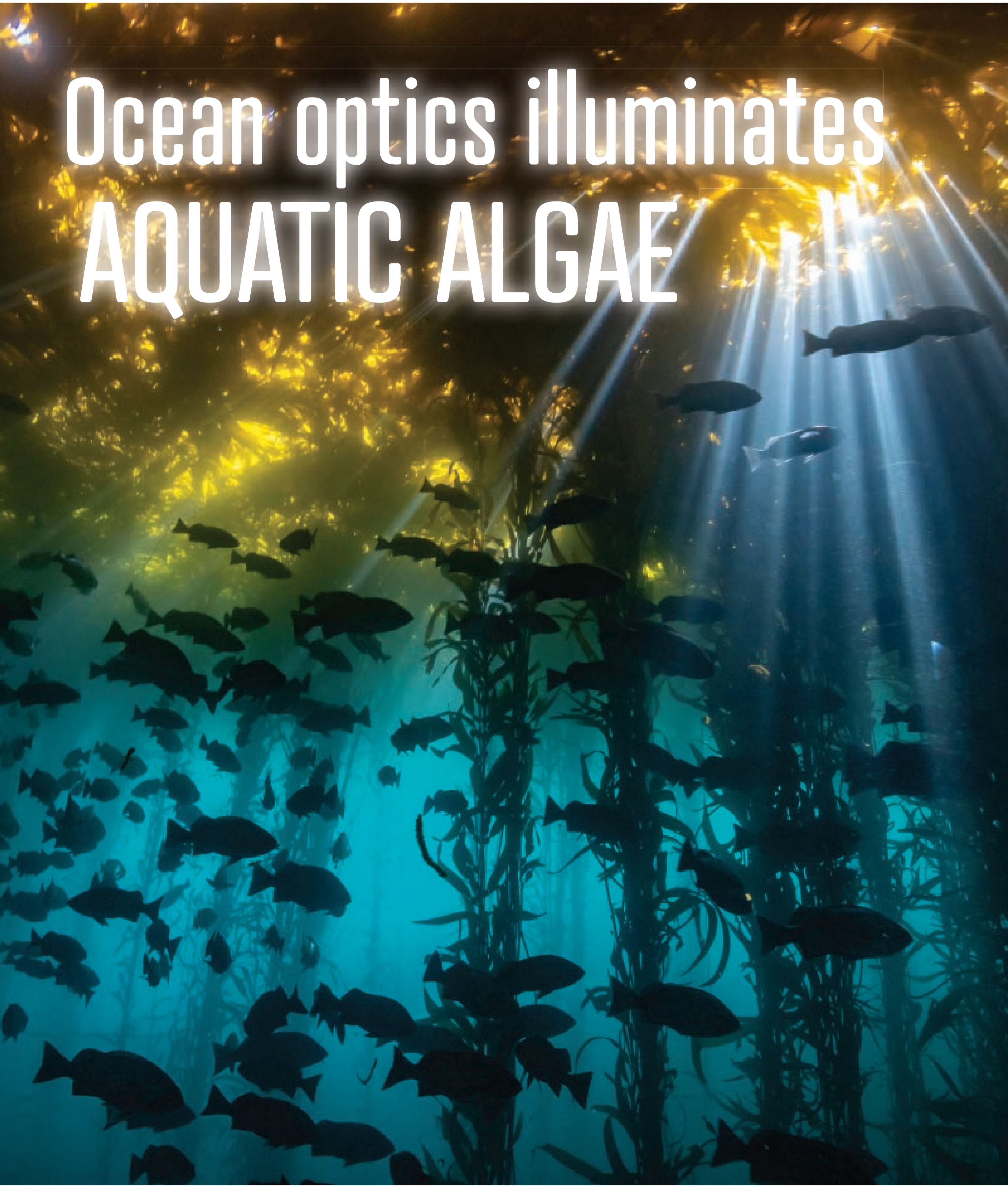
The IIJA established a financing mechanism within DOE to provide flexible, low-interest grants and loans to cover a portion of the cost of common-carrier CO_2 transport infrastructure development, lowering the risks for private-sector investment. The act allocates $2.1 billion over five years for those purposes.

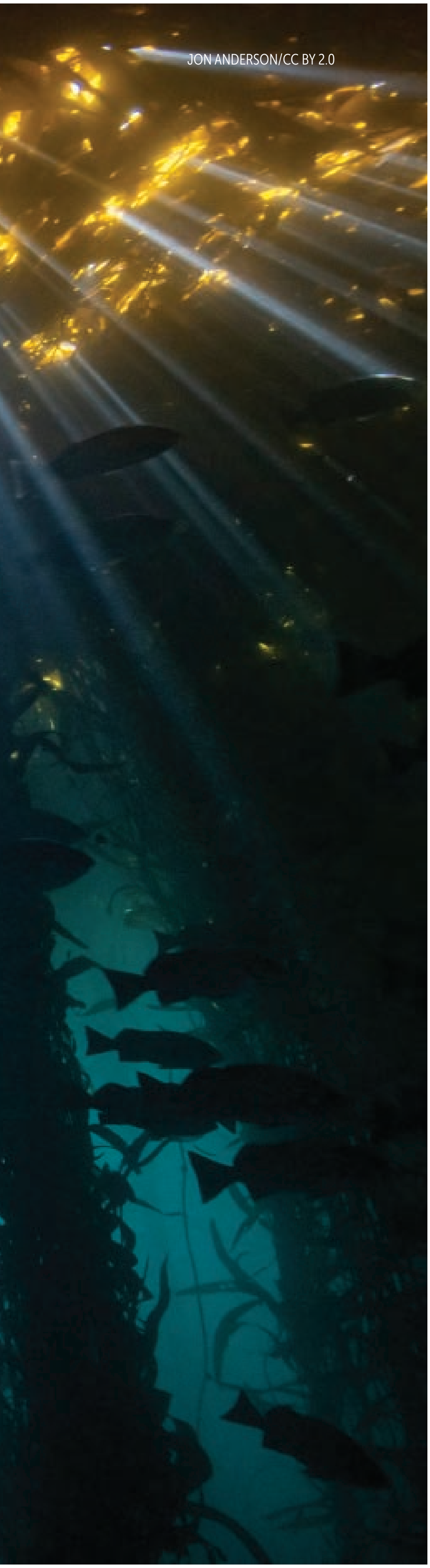
Currently no federal body has regulatory jurisdiction over the siting of interstate CO_2 pipelines. If authorized by Congress, the Federal Energy Regulatory Commission, which regulates interstate fossil fuel pipelines, or the Surface Transportation Board, which regulates freight rail, could be equipped for the role. But many large emitting states, including Texas and Louisiana, could transport from the emitting site to storage without having to cross state lines, Harrell notes.

The Department of Transportation's Pipeline and Hazardous Materials Safety Administration regulates CO_2 pipeline safety. It continues to update its regulations in the wake of a 2020 rupture of a CO_2 pipeline owned by Denbury Gulf Coast Pipelines in Mississippi. Although no one was injured, local authorities had to evacuate some nearby residents. Heavier than air, CO_2 disperses close to the ground and is an asphyxiant when concentrated.

David Kramer PT

Ocean optics illuminates AQUATIC ALGAE




JON ANDERSON/CC BY 2.0

Chuanmin Hu is a professor of oceanography at the University of South Florida in St Petersburg. He has over 30 years of experience in ocean optics and ocean color remote sensing.

Chuanmin Hu

Large masses of algae in the great Atlantic *Sargassum* belt and around the world affect local ecosystems and the environment. Satellite imagery, combined with traditional research techniques, is now helping scientists to study them.

Oceans and lakes are full of microscopic algal particles—single cells that are 1–200 μm in size and range in concentration from several tens to millions per liter of water. Various types of macroalgae, sometimes called seaweed, are also ubiquitous in oceans and lakes. *Sargassum*, *Ulva*, giant kelp, and others live either on the ocean floor or in surface water. Both microalgae and macroalgae are important to their local ecosystems. They also affect the environment around the world and the way humans interact with the ocean. Today, satellites and the developing field of ocean optics are rapidly changing the way scientists measure algae and their interactions with the environment.

Microalgal particles vary in size, shape, and type. The most common are cyanobacteria, often called blue-green algae, and diatoms, dinoflagellates, green algae, and coccolithophores (see figure 1), most of which can live in salty seawater, brackish coastal waters, and freshwater lakes. Knowledge of their distributions and concentrations is essential to understanding their individual ecological functions and how they respond to environmental changes.

Although microalgae, which are also called phytoplankton, are invisible to the human eye, the small plants still undergo photosynthesis and are the primary producers of carbon in the food chain. In fact, they are responsible for about half of the total primary production on Earth, with the other half provided by terrestrial vegetation.[1] Measuring microalgal particles, therefore, is critical to understanding their role in carbon sequestration, fisheries, and ecology.

At high concentrations, some microalgae form harmful algal blooms (HABs) that can produce toxins or deplete dissolved

FIGURE 1. MACROALGAE AND MICROALGAE. Clumps of the macroalgae *Sargassum fluitans III* **(a)** and *Sargassum natans I* **(b)** are abundant in the Atlantic Ocean. They are good habitats for aquatic life because their air bladders help them float on the ocean surface. (Images courtesy of Amy Siuda, Eckerd College.) **(c)** Common types of microalgae found in oceans and lakes. They are not drawn to scale or in natural colors; most are tens of micrometers in size. Each type may have hundreds of species that have different shapes than shown here. (Adapted from drawings and micrographs by Sally Bensusen, NASA EOS Project Science Office.)

oxygen. They kill fish and other animals, which causes alterations of the food web and other environmental effects. HABs can also disrupt the local economy. Cyanobacterial HABs have been recurrent in Lake Erie, Lake Okeechobee, and many lakes around the world. In the Gulf of Mexico, a single but long-lasting HAB of the toxic dinoflagellate *Karenia brevis* in 2018 cost Florida $317 million and 2900 jobs.[2] The repercussions harm not only industries that work in the ocean, such as commercial fishing and for-hire diving operations, but also such secondary industries as tourism and real estate because people stay away from affected regions.

Among the various types of macroalgae, pelagic (open-water) *Sargassum* in the Atlantic Ocean is of particular importance because it serves as a critical habitat for many marine animals. Atlantic *Sargassum* primarily has two species, *Sargassum fluitans* and *Sargassum natans* (see figure 1). Both reproduce vegetatively and live in the surface ocean until they die and sink to the ocean floor. *Sargassum* can cause problems when large amounts accumulate in coastal waters or on beaches. Early studies focused on its ecological functions and biomass through field measurements; recent advances in mapping *Sargassum* biomass from space serve as a prime example for how ocean optics addresses the needs of an emerging science topic.

Measuring microalgae in the field

Traditionally, microalgal particles are collected in the field and then analyzed under a microscope to characterize their size, shape, morphology, and taxa or species.[3] Field collection techniques have been extended to flow cytometry, in which pictures of microalgal cells suspended in a flow tube are captured on a camera and analyzed by either human analysts or computer programs. Recently, DNA barcoding has also been used to identify microalgal taxa and their living environment. Although the techniques are relatively accurate, they rely on physically handling the samples. Optical instruments, on the other hand, measure the microalgal particles through non-touch techniques.

The history of optical measurements of the ocean dates back to the initial efforts in the late 19th and early 20th centuries to understand the color of the ocean and how sunlight penetrates it. The light penetration, or water transparency, is determined by the depth at which a submerged Secchi disk—divided into quadrants and colored white and black—is no longer visible to the human eye. That is called the Secchi depth.[4] While relatively accurate, the handheld instruments measure only one property of the ocean water. To describe other optical properties, researchers needed to develop digital instruments that can accurately measure them.

Such requirements drove advances in ocean optics: the study of how light interacts with water and its constituents and how such interactions can be measured and interpreted. It employs radiative transfer theory, based on physical principles, to describe the optical properties of the ocean and its constituents and how light interacts with the atmospheric gases and particles (that is, aerosols).[5,6] A combination of physics and oceanography, ocean optics made its formal debut as an interdisciplinary science between the 1950s and the 1980s, when its pioneers systematically studied how natural light is absorbed, scattered, and fluoresced by marine particles and dissolved matter[7,8] (see figure 2).

Advances in ocean optics have led to new methods for measuring various microalgal particles. Early on, the optical properties (absorption and scattering) of microalgae and non-algae particles were measured in both the field and the laboratory using shining tubes, filter pads, and other instrumentation.[9] For example, the concentrations of chlorophyll *a*—the primary pigment used in photosynthesis—and other pigments were determined using fluorometers because once the microalgae absorb light at one wavelength, they can emit light at a longer wavelength.

Today, more sophisticated instruments and methods are available to measure all those properties. One such way is through high-performance liquid chromatography (HPLC). Different types of microalgae have different pigment compositions,[10] and each pigment has its own absorption spectrum, which is known *a priori*. The total pigment absorption spectrum can thus be decomposed into individual pigment absorption spectra from which the type and size of the microalgae can be estimated. For example, the freshwater cyanobacterium *Microcystis aeruginosa* (responsible for most freshwater HABs) can often be identified through its phycocyanin pigment, which contributes to its blue-green color, and the saltwater cyanobacterium *Trichodesmium* sp. can be identified through its phycourobilin, phycoerythrobilin, and phycocyanin pigments.

Alternatively, the ratio of microalgal backscattering to absorption can be used as an index to separate otherwise similar microalgae. That technique was employed to differentiate the toxic dinoflagellate *Karenia brevis* (responsible for most HABs producing red tides in the Gulf of Mexico) from diatoms because the former has a lower backscattering efficiency.[11] More importantly, surface spectral reflectance data from satellites have become a critical tool because inversion algorithms allow for the determination of absorption and scattering spectra.

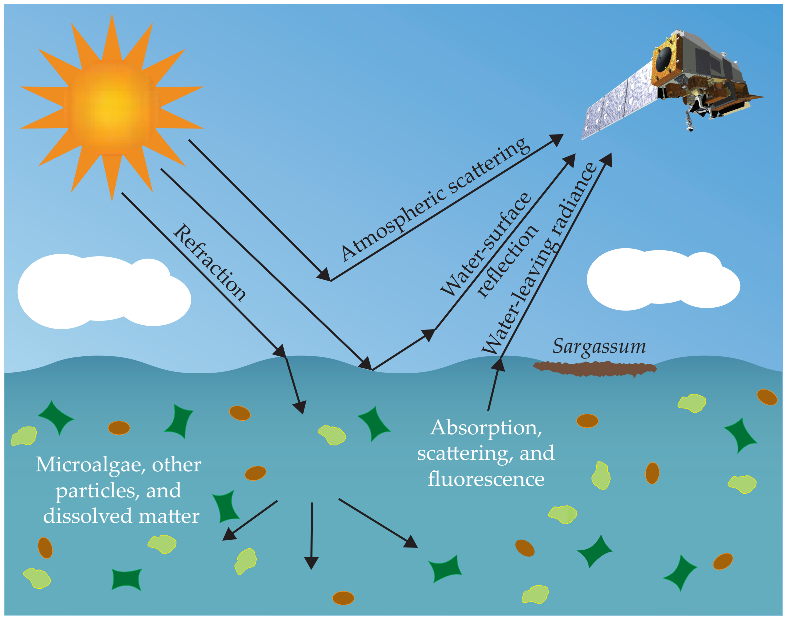


FIGURE 2. SUNLIGHT interacts with the atmosphere, the water, particles, and dissolved matter in the water before it is captured by a satellite. In remote sensing of ocean color, the atmosphere's contributions are corrected for, resulting in spectral surface reflectance that can be used to characterize microalgae and macroalgae both below and on the surface. (Image adapted from a diagram by Meng Qi; the *Joint Polar Satellite System-1* adapted from an illustration by Ball Aerospace.)

Measuring macroalgae in the field

Optical techniques are also used to measure macroalgae. For the case of the Atlantic pelagic *Sargassum*, most field and laboratory efforts rely on eye sighting, sample collection using a net,[12] and measuring the sample's pigment composition and chemical composition—carbon, nitrogen, phosphorus, arsenic, and more—using HPLC and other techniques. Optical measurements of *Sargassum* did not start until the 2010s.[13]

Because *Sargassum* on the surface forms individual clumps (see figures 1a and 1b) and multiclump mats, its reflectance is measured in the field in order to establish a relationship with its surface biomass density, or weight per area. The area of a *Sargassum* mat is determined by comparing its full size against a reference quadrat (see figure 3a) before collecting the marked-off *Sargassum* in a large net. It is weighed both wet, after removing the ambient water, and dry, after drying the sample in an oven. Around 80% of the wet weight is from the water alone.

The reflectance of the *Sargassum* mat is measured using a spectrometer with a fiber-optic probe, which collects the sunlight reflected in the quadrat. Figure 3b presents several sample reflectance spectra of *Sargassum* mats. The visible wavelength range 400–700 nm has a local reflectance maximum around 600–640 nm, which is why *Sargassum* appears yellowish or brownish. The local minimum at about 620 nm is caused by light absorption by chlorophyll *c* pigments, and the local minimum at around 670 nm is caused by light absorption by chlorophyll *a* pigments. In the near-IR wavelengths of 700–900 nm, the sharp increase is often called red-edge reflectance, which is typical for both terrestrial and large marine plants.

In addition to the reflectance and biomass density measurements in the field, *Sargassum* samples collected from the quadrat are also analyzed for their pigment composition and elemental concentrations of carbon, nitrogen, and phosphorus, among others.[13] Because the makeup of the samples is measured with the *Sargassum* reflectance, which can also be measured from satellites, field data are essential for validating satellite observations. Using data based on field and laboratory measurements, researchers develop algorithms that they apply to space-based measurements in order to estimate *Sargassum* biomass density and other properties.

With new ocean color remote sensing techniques, along with advances in theory, instrumentation, and algorithm development, scientists can not only measure and characterize the underwater light field but also observe global oceans and lakes from space. The satellite-detected signals are used to estimate surface spectral reflectance of the ocean through a process called atmospheric correction[6] (see figure 2). Reflectance carries information on pigment concentration, absorption, scattering, fluorescence, type, and size distribution of microalgal particles and on macroalgal biomass density and other properties. To derive those secondary properties, researchers rely on well-designed inversion algorithms developed from field and laboratory measurements. Because the whole process depends on interpreting the ocean's spectral reflectance, or

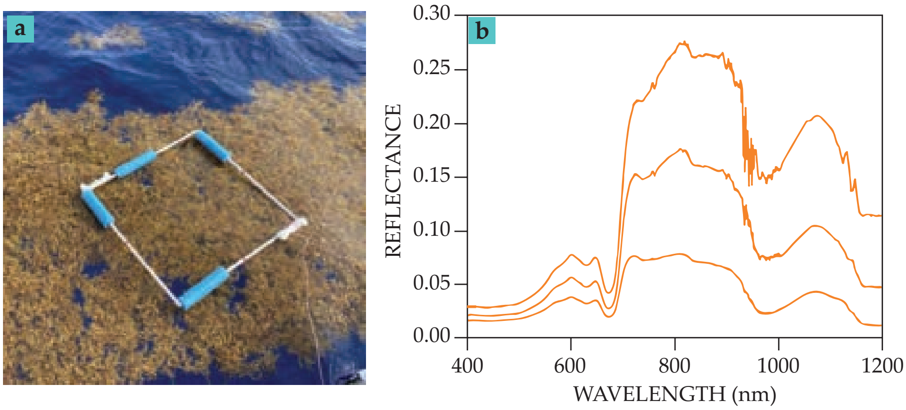


FIGURE 3. MEASURING *SARGASSUM* MATS in the Gulf of Mexico. **(a)** The biomass density is determined by weighing the collected *Sargassum* and using a reference quadrat, here 1 m on a side, to determine the size of the collected *Sargassum*. (Courtesy of Samuel Bunson.) **(b)** The mat's spectral reflectance is measured using a handheld spectrometer pointing at the mat. Differences in magnitude are primarily due to how much *Sargassum* (that is, biomass density in kilograms per square meter) is captured in the spectrometer's field of view. The relationship between reflectance and density is used to develop algorithms to calibrate and validate space-based measurements, which provide only reflectance data.

color, it is also called ocean color remote sensing. (The same terminology is also used for freshwater lakes.)

The history of ocean color remote sensing goes back to the pioneering works of many individuals.[14] Since 1978, when the proof-of-concept Coastal Zone Color Scanner was one of eight instruments aboard the *Nimbus-7* spacecraft, modern satellite sensors specifically designed to measure reflected sunlight have been equipped with more spectral bands, better signal-to-noise ratios, and finer spatial resolutions. Recent missions also allow for more frequent revisits to the same area. The satellite-derived surface reflectance measurements produce unprecedented data to study not only light penetration in the ocean but also, through the use of inversion algorithms, properties of microalgal particles and macroalgal mats.

Indeed, since the 1980s the number of publications that use the phrases "ocean optics" and "ocean color" has grown exponentially. Global distribution maps of microalgal particles that show concentration, absorption, scattering, type, size, and other variables have been made available by various space agencies (including NASA, at https://oceancolor.gsfc.nasa.gov). Those data products have been used by the research community to not only quantify chlorophyll *a* and primary production but also differentiate different phytoplankton functional groups (see figure 4), monitor HABs, and understand their physiological response to the changing ocean environment. For example, US federal and state agencies have applied tailored algorithms to multisensor ocean color data to generate and distribute HAB bulletins for HABs in both oceans and lakes.

Measuring *Sargassum* from space

Ocean color remote sensing is now a crucial tool for studying not only microalgae but also macroalgae. Among the many types of macroalgae, the Atlantic pelagic *Sargassum* represents a special case because the recurrent great Atlantic *Sargassum* belt (GASB)[15] represents an emerging phenomenon that has significant implications for research, the environment, tourism, and the economy.

The relationship between *Sargassum*'s reflectance and biomass density—established from field measurements—can be applied to satellite-measured reflectance in each valid image pixel to map its biomass density from space. In that process, contributions from the atmosphere and other perturbation factors are removed (see figure 2), and radiative transfer equations estimate how such removals may modulate the functional relationship. Cloud cover and other factors frequently lead to data gaps in images. But gaps in a given location during a certain period can be filled by calculating an average *Sargassum* biomass density from all valid pixels in the images of preceding and subsequent days.

The distribution map of *Sargassum* density for the month of June 2021 is shown in figure 5, where a continuous *Sargassum*-heavy "belt," the GASB, is observed from western Africa to the Gulf of Mexico. The water area in the belt is about 5 × 10^6 km^2,

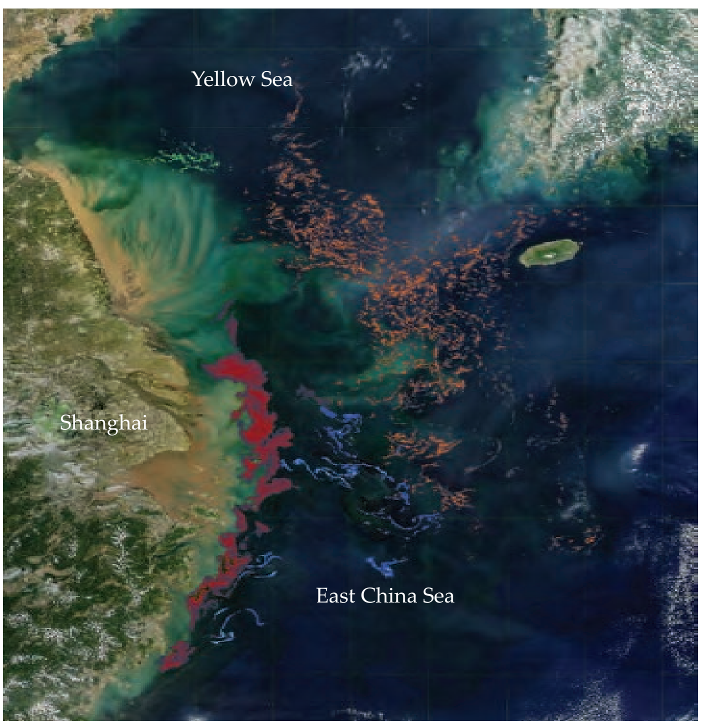


FIGURE 4. A BLOOMING OCEAN of algae captured by NASA's Moderate Resolution Imaging Spectroradiometer on the *Aqua* satellite over the East China Sea and Yellow Sea in 2017. Among the detected ocean algae are the microalgae *Prorocentrum donghaiense* (red) and *Noctiluca scintillans* (blue) and the macroalgae *Sargassum horneri* (orange) and *Ulva prolifera* (green). (Satellite image courtesy of Norman Kuring, NASA's Goddard Space Flight Center; additional algae traces highlighted by Lin Qi.)

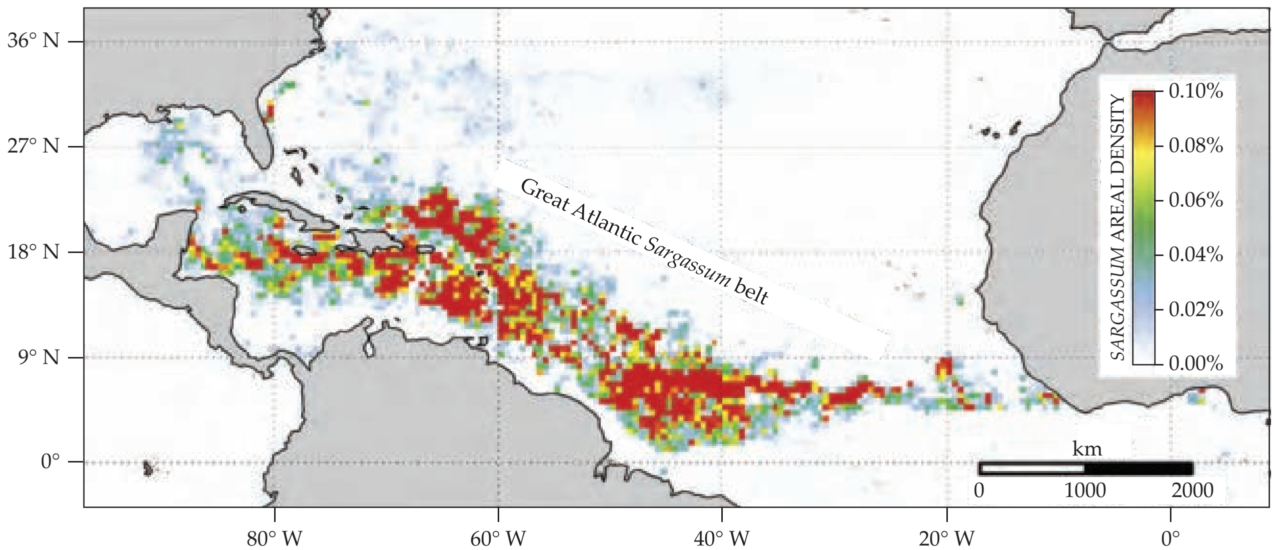


FIGURE 5. THE GREAT ATLANTIC *SARGASSUM* BELT in June 2021, which spanned an area of about 5×10^6 km^2. The *Sargassum* itself was estimated to be about 6000 km^2 when aggregated together, with a total wet biomass of 18 million metric tons (t). The areal density of 0.1% corresponds to biomass density of 3.3 t/km^2. (Data courtesy of MODIS.)

and the *Sargassum* mats scattered around would occupy a surface area of about 6000 km^2 if aggregated together, with an estimated total wet biomass of 18 million metric tons.

The discovery of the GASB, which has been recurring annually since 2011, resulted from advances in ocean optics and ocean color remote sensing. Traditionally, *Sargassum* was known to be abundant in the Sargasso Sea—a subocean in the North Atlantic named after the macroalgae—but satellite data show much higher abundances of the macroalgae in the tropical Atlantic in recent years. Researchers are hard at work studying the GASB, including how it was formed, whether it indicates a regime shift resulting from a changing climate, how it affects local biology and ecology, how it may change carbon cycling, and whether it can be used to sink carbon to the ocean floor.

Sargassum is a critical habitat for ocean life, but excessive amounts of it in coastal waters and on beaches have caused tremendous problems because dead *Sargassum* can sink to the bottom and smother corals and seagrasses. Rotten *Sargassum* on beaches can destroy turtle nests, attract insects, grow and spread bacteria, and cause respiratory problems in humans. All of those issues adversely affect local tourism and the economy, turning *Sargassum* from a critical habitat to a beach nuisance.

Countering *Sargassum*'s negatives will require more research efforts in understanding its biology and forecasting its blooms and in developing new methods to mitigate its effects and make it into useful products, such as fertilizers. Ocean optics and ocean color remote sensing are expected to continue playing an essential role in all those aspects. Improved understanding of the *Sargassum* biomass, growth rate, and changing patterns will provide more accurate data for forecasting models and for improved monitoring and tracking capacity through the application of artificial intelligence.[16]

Globally, *Sargassum* is not the only macroalgae that can be detected, characterized, and quantified using optical means. Other types, either floating on the water surface or growing on the shallow bottom, have also been reported worldwide. They include *Ulva prolifera*, a green macroalgae, in the Yellow Sea; *Sargassum horneri*, a brown macroalgae that grows on the shallow ocean floor but can be detached to float on the surface, in the East China Sea[17] (see figure 4); giant kelp in California (see the opening image); and seagrass around the world's coastal oceans. In each case, the macroalgae can be studied both in the field and from space, based on the same optical principles applied to the Atlantic *Sargassum*. With more research dedicated to macroalgae, their roles in carbon science and other Earth-science disciplines will be better recognized and quantified.

Satellites big and small

After a century of development, ocean optics continues to greatly affect technology and science. The theory of underwater visibility has been revised from relying on assumptions to enabling direct estimates of water clarity. New sensors have been invented to measure variable fluorescence. New flow cytometers have been created to analyze thousands of microalgal cells per second. Those and other developments allow for more accurate assessments of phytoplankton physiology, and some sensors have been customized to be mounted on autonomous platforms, such as ocean gliders and the floats used by the international Argo program to collect ocean data.

The sensors include direct and derivative measurements of light scattering, absorption, fluorescence, attenuation, and reflectance—thus making it possible to characterize the underwater light environment and to automatically estimate microalgae and other ocean constituents. For example, as it drifts with ocean currents, each Argo float cycles between the surface and depths of 2000 m every 10 days, collecting a depth-dependent profile of ocean properties through its mounted sensors. There are 4000-plus Argo floats in the world's oceans, so such bio-optical data will provide an unprecedented capability for mapping microalgae and other ocean properties in three dimensions, developing remote-sensing algorithms, and validating remote-sensing data products.

Those developments also support new polar-orbiting and geostationary satellite missions that use detectors that combine spectroscopy with imaging, polarimetry, and active techniques, such as lidar. Planned efforts include NASA's forthcoming Plankton, Aerosol, Cloud, Ocean Ecosystem (PACE) mission; NASA's Geostationary Littoral Imaging and Monitoring Radiometer (GLIMR), currently being developed; and NOAA's future Geostationary Extended Observations (GeoXO) satellite system.

Of particular importance are the emerging constellations of easy-to-produce small satellites known as CubeSats that can measure land and coastal waters at a spatial resolution of just a few meters with near-daily revisits around the globe. Such a capacity has never been possible until recently because traditional satellites could not address the simultaneous requirements of high spatial and high temporal resolutions needed for studying the fast-changing microalgae and macroalgae with fine details. Indeed, the PlanetScope constellation of more than 200 CubeSats already has shown a much-improved capacity to track macroalgae near beaches.[18]

Computer artificial intelligence has found increased applications in field and laboratory measurements and optical remote sensing, often with more robust algorithms and data products than what's offered by traditional methods. All the advances will continue to enhance our capability to characterize, quantify, and understand both microalgae and macroalgae in the field and from space.

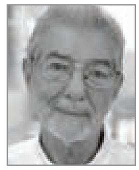
REFERENCES

1. A. Morel, D. Antoine, *Science* **296**, 1980 (2002).
2. J.-P. Ferreira et al., *Tour. Econ.* **29**, 812 (2023).
3. K. A. Steidinger, M. E. Meave del Castillo, eds., *Guide to the Identification of Harmful Microalgae in the Gulf of Mexico*, vols. 1 and 2, Florida Fish and Wildlife Conservation Commission (2018).
4. P. A. Secchi, *Nuovo Cimento* **20**, 205 (1864).
5. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*, Academic Press (1994).
6. H. R. Gordon, *Physical Principles of Ocean Color Remote Sensing*, University of Miami (2019).
7. N. G. Jerlov, *Marine Optics*, Elsevier (1976).
8. R. W. Preisendorfer, *Hydrologic Optics*, vols. 1–6, NOAA (1976).
9. C. S. Yentsch, *Limnol. Oceanogr.* **7**, 207 (1962).
10. A. Bricaud et al., *J. Geophys. Res. Oceans* **109**, C11010 (2004).
11. J. P. Cannizzaro et al., *Cont. Shelf Res.* **28**, 137 (2008).
12. D. S. Goodwin, A. N. S. Siuda, J. M. Schell, *PeerJ* **10**, e14079 (2022).
13. M. Wang et al., *Geophys. Res. Lett.* **45**, 12359 (2018).
14. J. Acker, *The Color of the Atmosphere with the Ocean Below: A History of NASA's Ocean Color Missions*, CreateSpace Independent Publishing Platform (2015).
15. M. Wang et al., *Science* **365**, 83 (2019).
16. C. Hu et al., *Remote Sens. Environ.* **289**, 113515 (2023).
17. L. Qi et al., *Geophys. Res. Lett.* **49**, e2022GL098185 (2022).
18. S. Zhang et al., *IEEE Geosci. Remote Sens. Lett.* **19**, 1503605 (2022).

ICE FRACTURING

PETER REJCEK/NSF

Erland Schulson is director of the Ice Research Laboratory and the George Austin Colligan Distinguished Professor of Engineering at Dartmouth College in Hanover, New Hampshire.



Erland M. Schulson

The process comprises the nucleation, propagation, and interaction of cracks. Understanding their micromechanics in ice is likely to help scientists understand fracture in all kinds of materials.

Growing across the floating shelf of Thwaites Glacier in West Antarctica is a giant crack that will eventually become long enough to cause the ice shelf to collapse. That will mean the loss of a buttress against the advance of ice upstream of where the glacier is grounded to bedrock. And as a result, sea level will rise at an increasing rate. Thwaites is not alone. Distributed around some 70% of the continent's perimeter are many such shelves that perform the same function. They, too, are vulnerable to collapse.

The breakup of ice in the Arctic has similar environmental consequences. Although it is decreasing in extent because of global warming, sea ice covered an area of more than 14 million km^2 in March of this year. That's greater than the area of the continental US and Mexico combined. The ice cover serves as a barrier to the transfer of heat and moisture from the ocean to the atmosphere. But it is riddled with defects. Under the action of wind and ocean current, cracks form and open, accelerating that transfer.

In those and other instances, the loss of the ice's structural integrity is rooted in the nucleation, propagation, and interaction of cracks. Exactly how that happens in large masses of ice is an open question and remains a topic of current research. Reviewing the various hypotheses is beyond the scope of this article. My purpose, rather, is to describe the physical mechanisms that govern fracture and, where possible, to elucidate the larger picture.

Scientists have gained insight into those mechanisms largely through experiments in the laboratory, where the main factors at play—temperature, rate of loading, stress, and microstructure—can be controlled and varied in a systematic manner and their effects quantified. The combination of optical transparency and millimeter-scale microstructure in ice allows cracks inside the material to be visible to the unaided eye, which helps in analyzing its fracture mechanics.

This article is limited to the fracture of ice formed from water and loaded under terrestrial conditions. For citations to the original literature and derivations of the functional relationships discussed here, I recommend consulting references 1 and 2.

Structure of ice

The microstructure of ice, like that of any material, is key to understanding its behavior. Whether in the form of a polar glacier, a floating ice cover, or an ice cube in the refrigerator, ice is made from the aggregation of grains joined at their boundaries. When formed under terrestrial conditions, the grains are typically 1–10 mm in diameter. A grain's shape may be either equiaxed, in which it has roughly the same length in any direction, or columnar, in which it is longer in one direction than the others, much like a candle, as shown in figure 1.

Equiaxed aggregates, termed granular ice, describe glaciers and icebergs and form through the consolidation of snow. Columnar-shaped aggregates, which include a small volume fraction (0.05, typically) of submillimeter-sized pockets of brine, characterize floating sea ice covers and are produced through unidirectional solidification—the vertical transfer of heat along the columns. Regardless of their shape, the grains adopt a crystalline structure in which the H_2O molecules are arranged three dimensionally in a periodically repeating array and are held together by hydrogen bonds.

The basic building block, or unit cell, possesses hexagonal symmetry, denoted *Ih*. Although the orientation of the unit cell remains fixed in any given crystal, in granular ice it varies from grain to grain in an apparently random manner. So even though hexagonal symmetry imparts anisotropic behavior to an individual crystal—making each crystal's properties dependent on direction—the aggregate as a whole exhibits isotropic behavior.

In columnar ice, on the other hand, the unit cell adopts a preferred orientation. Depending on the conditions of solidification, the long axis of the cell, termed the *c*-axis, either aligns with the vertical direction, running parallel to the long axis of the grains, or is confined to the horizontal plane, running generally in various directions within that plane. As a result,

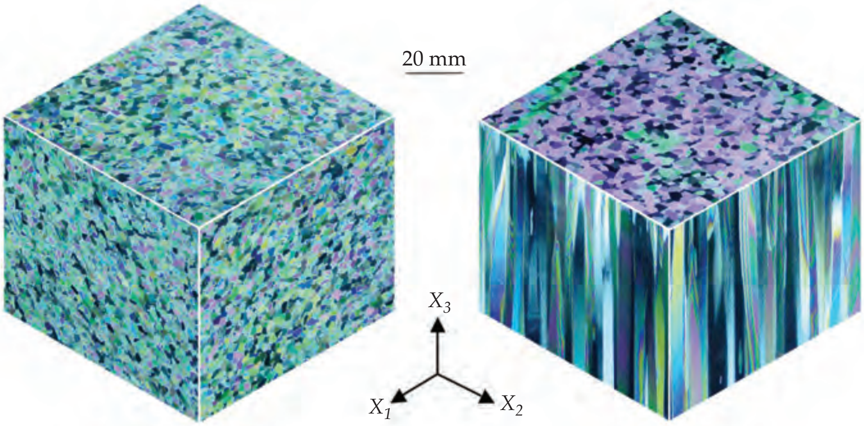


FIGURE 1. GRANULAR ICE (left) and columnar ice (right), photographed under polarized light. The granular ice was produced through the consolidation of millimeter-sized grains of ice. The columnar ice was produced through the unidirectional crystallization of fresh water along the vertical direction X_3. (Courtesy of Narayana Golding and Daniel Iliescu, Dartmouth College.)

columnar aggregates possess a preferred orientation and exhibit anisotropic behavior.

Inelastic behavior

Ice does not always break up or fracture when loaded—that is, behave in a brittle manner. When loaded slowly, it deforms in a plastic or ductile manner. Antarctic glaciers, for instance, flow under the force of gravity and can develop strain in excess of unity (100%) without breaking. Similarly, icicles can bend through an angle up to 75° under the force of wind. That kind of behavior originates through the stress-driven slipping of the individual grains on the basal planes of the *Ih* lattice. The process is analogous to the slipping of cards in a deck. It involves the movement, or "glide," of line defects, termed crystal dislocations, and begins once the shear stress on the planes reaches a critical level.

Once activated, slip accommodates the strain ε imparted by the applied load, and the aggregate deforms plastically. The plastic strain rate $\dot{\varepsilon}$ increases nonlinearly with the applied stress σ and is described by the equation $\dot{\varepsilon} = B\sigma^n$. The parameter B is a temperature-dependent materials constant and a measure of the ease of plastic flow. The stress exponent, although the subject of debate, has often been taken to have the value $n = 3$.

The equation dictates how the flow stress increases with rising strain rate, scaling as $\sigma_D \sim \dot{\varepsilon}^{1/n}$, in which the subscript D implies ductile behavior. At a certain rate, the stress required for plastic flow becomes so high that any cracks—whether they were present initially or were nucleated during loading—will begin to lengthen, or propagate. When they do elongate, the behavior of the aggregate changes from ductile (D) to brittle (B).

Microstructure enters the picture in two ways. First, the plastic parameter B depends on crystallographic texture, the material's preferred orientation. For a floating cover of columnar-grained ice loaded across the columns, B is lower for aggregates in which the *c*-axis of the unit cell runs parallel to the columns—that is, oriented in the vertical direction.

The lower value means that the bending load for plastic flow at a strain rate imposed by the rate of loading is greater than the bending load for aggregates in which the *c*-axes are oriented in the horizontal plane. Given that the resistance to crack propagation is relatively insensitive to *c*-axis orientation, the effect of the vertical orientation is to reduce the strain rate that marks the transition from ductile to brittle behavior.

The second way microstructure appears is through grain size. For reasons described below, the load required to propagate cracks decreases as the grain size increases and is lower under tension than under compression. Thus, because of the influence of microstructure and stress, the strain rate that marks the DB transition, expressed as $\dot{\varepsilon}_{DB}$, is not a constant but decreases as both the ease of plastic flow and the brittle failure stress decrease. Typical values of the transition strain rate for pristine ice—that is, material initially free of cracks—are $\dot{\varepsilon}_{DB} = 10^{-8}$ to 10^{-7} s^{-1} under tension and $\dot{\varepsilon}_{DB} = 10^{-4}$ to 10^{-3} s^{-1} under compression.

Brittle failure

When activated by the appropriate combination of factors, brittle failure is marked by a sudden loss of load-bearing ability. At play is crack mechanics, particularly its propagation. Cracks weaken a material by concentrating stress in proportion to the square root of their lengths. The effect of those cracks on strength is expressed in terms of a crack-tip stress intensity factor K, defined as the product of the crack's length $\sqrt{l}$ and the tensile stress σ_T, which acts normal to the plane of the crack; that is, $K = Y\sigma_T\sqrt{\pi l}$, where Y is a geometrical factor of order unity.

Because of their thermomechanical history, natural icy bodies generally contain cracks whose lengths can reach more than a meter, whereas pristine ice—the kind most often studied in the laboratory—develops cracks whose lengths are set by the grain size of each aggregate. For both natural and pristine ice, the brittle-failure stress σ_B increases with decreasing grain or crack size and may be described by the relationship $\sigma_B = CK_C l^{-0.5}$, where C is a dimensionless constant that incorporates the applied state of stress (either tensile, compressive, or multiaxial) and K_C denotes a critical level of the stress intensity factor.

The parameter K_C is a materials property termed fracture toughness and is a measure of resistance to crack propagation. On Earth's surface, the fracture toughness of ice is relatively insensitive to temperature and to microstructure. It has a low value of around 100 kPa·m$^{0.5}$ for both freshwater ice and sea ice.[3] (Note that 100 kPa is equivalent to one atmosphere of pressure.) The low toughness originates in the low strength of the intermolecular hydrogen bonds that hold the *Ih* crystal together.

Compared with other natural materials, ice offers little resistance to propagating cracks. The combination of low toughness and crack length generally greater than 1 mm dictates that pristine ice undergo brittle failure at quite low stress—on the order of about 1 MPa under tension and 10 MPa under compression. Defective natural ice is even weaker because it has larger cracks. Even so, when a large mass, such as a sheet of sea ice, under the action of wind pushes against an offshore structure, such as the lighthouse shown in figure 2, the combination of material strength and contact area can induce a force that is greater than the maximum induced by ocean waves over a period of 100 years[4] and cause the structure to fail.

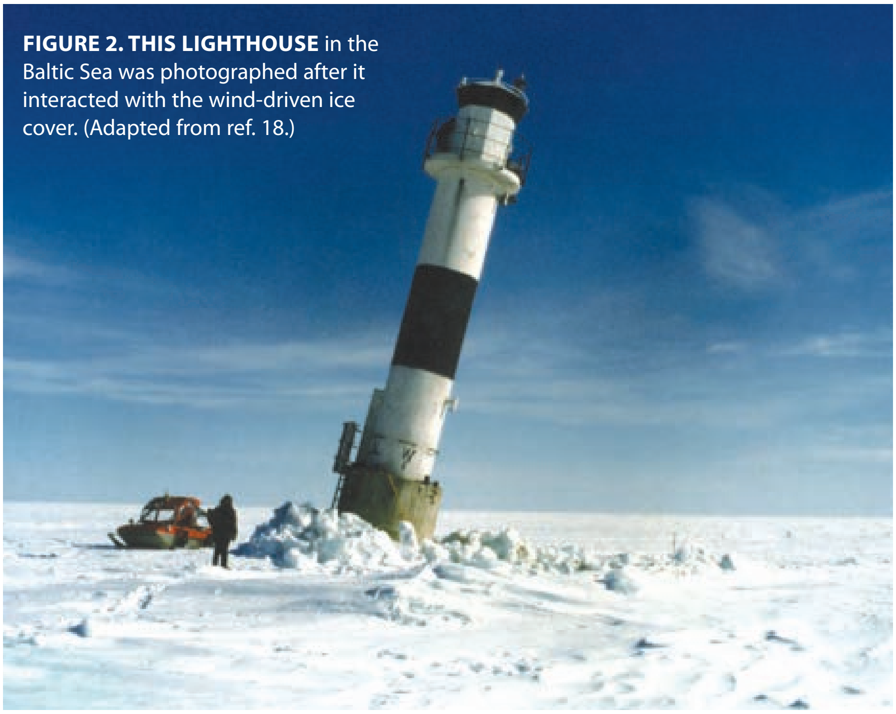


FIGURE 2. THIS LIGHTHOUSE in the Baltic Sea was photographed after it interacted with the wind-driven ice cover. (Adapted from ref. 18.)

To appreciate why ice fails at a higher stress when loaded under compression than under tension, consider the underlying mechanisms. In pristine material, cracks must first nucleate. That happens as stress builds up locally within the body, either by dislocations impinging on grain boundaries and piling up there or by grains sliding along their boundaries. Once the grain-boundary stress is high enough, it is relieved by the creation of grain-sized microcracks, inclined at around 45° to the direction of loading. Tension acts to open those microcracks and is usually great enough when combined with a long enough crack to generate a sufficiently large stress intensity factor and cause unstable crack propagation. And that unstable propagation may happen either when the crack first nucleates or shortly thereafter, when it creates a fracture plane perpendicular to the direction of loading.

Compression, on the other hand, closes cracks, which leads to a more complicated failure process. Because of their inclination, the upper crack faces tend to slide over the lower faces as long as the shear stress on the plane of the crack is sufficient to overcome friction. At the crack tips, sliding generates tension, which is relieved through the initiation of short, out-of-plane extensile or secondary cracks, known as wing cracks, that tend to align along the direction of loading.[5,6]

When those wing cracks are produced, tension is transferred to their tips, thereby accounting for how a compressive load can generate an internal tensile stress. As the load increases, sliding continues: The mouths of the wing cracks open, the tensile stress at their tips increases, and the wings begin to lengthen in a stable manner once the crack-tip stress intensity factor reaches the critical level. The wing cracks interact with other wings, the ensemble of which eventually forms a macroscopic fault that runs parallel to the direction of loading. At that point, terminal failure ensues, and the ice loses its load-bearing ability. Frictional sliding, in other words, is the reason why compressive strength is greater than tensile strength.

Natural icy bodies are seldom loaded simply by pushing or pulling in one direction. More commonly, the ice is loaded three dimensionally—for example, by bending about two axes or through confinement, as in a pressure cell or beneath an indenter. Multiaxial loading has little effect on the tensile stress required to cause a primary crack to propagate, and so it has little effect on tensile strength. On compressive failure, however, the effect is large. That's evident from the fact that a confining pressure as small as 10% of the applied compressive stress almost doubles the strength. The reason is twofold: Confinement acts to close wing cracks and therefore reduces the stress intensity at their tips. It also raises the applied stress acting normal to the plane of the primary crack. And through friction, that component of stress increases the resistance to sliding.

The frictional model just described implies that sliding occurs uniformly across the opposing faces of the parent cracks. Confinement, however, induces nonuniform sliding. That has the effect of generating on one side of the parent crack (and sometimes on alternating sides) localized zones of tensile stress. The tension is again relieved by cracking, in this case by the initiation of sets of other wing cracks. Each set creates, in effect, microcolumns that are fixed on one end and free on the other. Frictional drag across their free ends bends and then breaks the microcolumns under an applied compressive load lower than that required to buckle the columns.

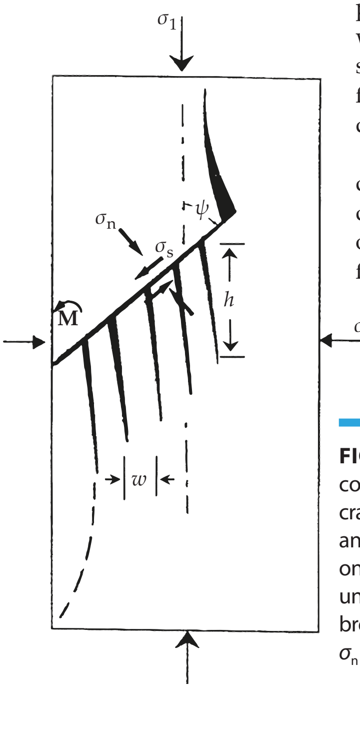


FIGURE 3. COMB CRACKS, in a schematic. Fundamental to the brittle failure of ice loaded under multiaxial compressive stresses σ_1 and σ_3 is frictional sliding across a primary crack. In the piece of ice sketched here, wing cracks—each of height *h* and separated by a width *w*—emerge from one side of the primary crack, inclined at an angle ψ to the direction of loading. Those wing cracks create microplates that resemble the teeth in a comb, fixed on one end and free on the other. Frictional drag across the free ends of the microplates produces a moment **M** per unit depth. The moment is more important than the axial load and eventually becomes large enough to bend and break the microplates and trigger a macroscopic shear fault. The normal and shear components of the stress tensors, σ_n and σ_s, are pictured. (Adapted from E. M. Schulson, D. Iliescu, C. E. Renshaw, *J. Geophys. Res.* **104**, 695, 1999.)

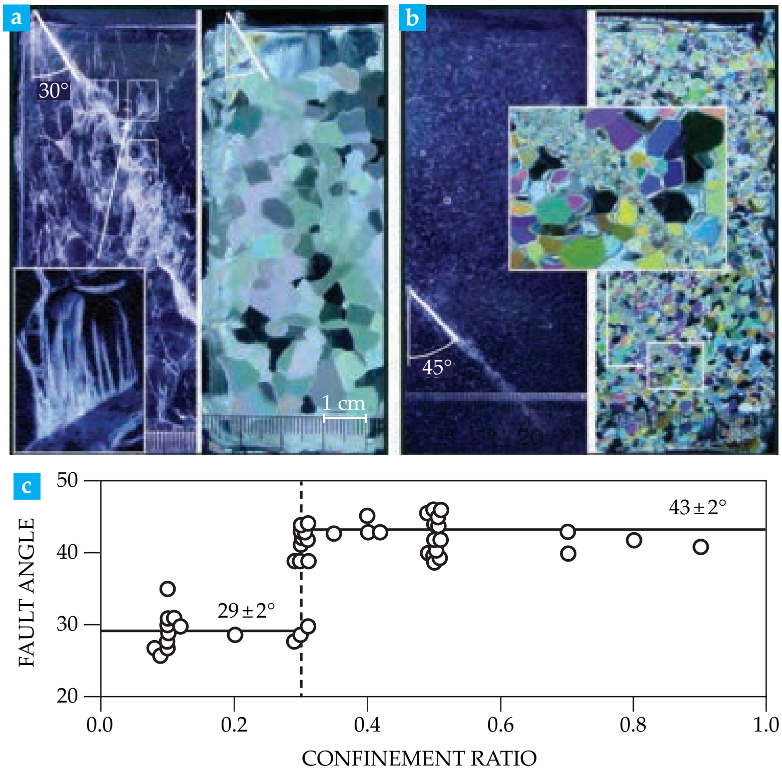


FIGURE 4. SHEAR FAULTS and plastic faults under confinement. **(a)** This thin (1 mm) section of a coulombic shear fault was formed in freshwater ice compressed at −10 °C under little ($R = 0.2$) three-dimensional confinement—the ratio of the least compressive stress to the most. It was then photographed under natural (left) and polarized (right) light. A comb crack (inset) emerges from one side of a primary crack. **(b)** A thin section of a plastic fault was formed in the same kind of ice under a high degree ($R = 0.5$) of confinement and photographed under natural (left) and polarized (right) light. Recrystallized grains along the fault are shown in the inset. The plastic fault is more steeply inclined (45°) than the coulombic fault (30°) to the direction of shortening. **(c)** The fault angles are measured with respect to the direction of shortening and are plotted as a function of the confinement ratio. Frictional sliding is suppressed above $R = 0.3$. (Adapted from ref. 2.)

The mechanism acts like the sliding of a thumb across the teeth of a comb—hence the term comb–crack mechanism, pictured schematically in figure 3. It changes the orientation of the fault from being parallel to the direction of shortening under no confinement to being inclined about 30° to that direction. (The angle is set by the coefficient of friction μ, through the relationship $\tan 2\theta = \mu^{-1}$.) The $\theta = 30°$ angle implies a coefficient of $\mu = 0.6$, in agreement with an independent measurement of the friction. Faults so oriented are called coulombic faults in recognition of the important role that frictional sliding has in their development.[2]

Satellite imagery has revealed intersecting sets of strike–slip-like faults, termed linear kinematic features,[7–9] that run hundreds of kilometers through the winter sea-ice cover on the Arctic Ocean. Those features were activated intermittently through wind-induced compressive stresses on the order of kilopascals. And they closely resemble intersecting coulombic shear faults generated in the laboratory from winter sea ice under megapascal compressive stresses (see reference 1, chapter 15).

The three orders of magnitude difference in stress reflects a six orders of magnitude difference in the length of sliding, primary cracks—kilometer versus millimeter—and the square-root dependence of the stress intensity factor on crack length. The coefficient of friction of ice on ice is independent of scale and would account for the similarity in angle of intersection on both large and small scales.

In 2001 Carl Renshaw and I developed a model based on the comb–crack mechanism.[10] It incorporates only physically measurable parameters—fracture toughness, crack size, the coefficient of kinetic friction, and the ratio of the confining stress to the major stress. It captures the effect of confinement on the brittle compressive strength of ice. And in using no adjustable parameters, the model also accounts for the strength of various rocks and minerals.

Plastic versus coulombic faulting

So far in this article, crack mechanics has accounted for the sudden loss of load-bearing ability by ice on rapid loading. That is not always the case, however. Under compression, the hydrostatic component of the stress tensor becomes large enough to suppress frictional sliding, and it thereby shuts down the comb–crack mechanism and prevents failure via coulombic faulting.

Yet the ice continues to deform. The shear component of the stress tensor reactivates dislocation glide—the movement of the dislocations across crystallographic planes—accompanied at high strain rates by adiabatic heating. That dislocation glide eventually leads to failure through the development of a different kind of fault known as a plastic fault. The transition from coulombic faulting to plastic faulting occurs when the degree of confinement, defined as the ratio of the least compressive stress to the most, reaches a critical level R_{cp}, which is set by the coefficient of friction: $R_{cp} = [(\mu^2 + 1)^{0.5} + \mu]^{-2} = 0.3$ for $\mu = 0.6$.

Plastic deformation generates heat. Indeed, about 80% of plastic work is expended as heat. Although ice is a poor thermal conductor, at low strain rates the heat generated through dislocation glide is dissipated over time on selective planes of the *Ih* crystals. At high strain rates, on the other hand, there's not enough time for the heat to dissipate. As a result, the slipping zones simply heat up and become weaker. (Recall that the parameter B increases with temperature.) Thermal softening dominates strain hardening, and plastic strain remains confined to narrow bands, within which the combination of higher temperature and high strain triggers a solid-state phase transformation called dynamic recrystallization.

The result is a band composed of soft, tiny grains, oriented at around 45° to the direction of shortening. What's thus formed on a plane of maximum shear stress is a plastic fault (see, for example, reference 11). Figure 4 shows a plastic fault in granular ice and, for comparison, a coulombic fault in the same kind of material under moderate confinement.

In practice, objects can experience a lot of compression when one dents another, as would happen, for instance, when an iceberg and a ship collide. Indentation experiments on both large and small bodies revealed that ice can fail at strengths as high as 50 MPa. The near-surface, highly confined regions are essentially crack-free and composed of recrystallized grains,[12,13] indicative of failure via plastic faulting.

Like coulombic faulting, plastic faulting is not limited to ice. Evidence exists that Earth minerals antigorite, granite, and olivine fail by the same mechanism when rapidly loaded under a high degree of triaxial confinement.[14]

Returning to the Antarctic ice shelf that I began this article with, crack growth is a slow process, at least up to the point of collapse. Although several mechanisms are at play, including surface flooding and hydrofracturing, the growth might be triggered by ocean swells.[15,16] If so, then cyclic variations in the stress intensity factor ΔK may be important, possibly attenuated by crack-tip blunting through either creep—that is, time-dependent plastic flow—or impingement on soft or heterogeneous regions in the shelves.

Perhaps crack growth and its time of collapse could be modeled according to the Paris–Erdogan law,[17] which states that the increase in crack length per loading cycle scales with the change in the stress intensity factor—that is, with the product of the change in tensile stress and the square root of crack length. But so far as I'm aware, no analysis of that kind has yet been made.

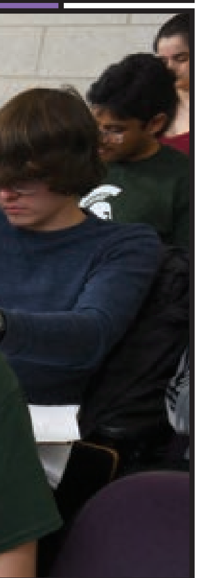
I thank colleagues Harold Frost and Carl Renshaw for fruitful discussions over the years and my students, who performed the research on which some of this article is based. I am grateful for financial support from NSF.

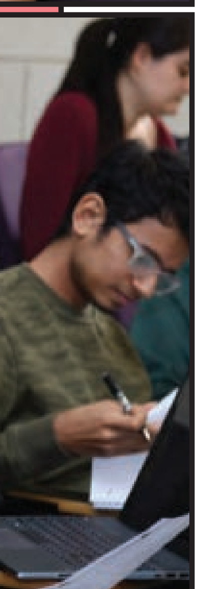
REFERENCES

1. E. M. Schulson, P. Duval, *Creep and Fracture of Ice*, Cambridge U. Press (2009).
2. E. M. Schulson, C. E. Renshaw, *Annu. Rev. Earth Planet. Sci.* **50**, 323 (2022).
3. K. L. Litwin et al., *J. Geophys. Res.* **117**, E08013 (2012).
4. American Petroleum Institute, *Recommended Practice for Planning, Designing and Constructing Structures and Pipelines for Arctic Conditions*, 2nd ed. (1995).
5. W. F. Brace, E. G. Bombolakis, *J. Geophys. Res.* **68**, 3709 (1963).
6. M. F. Ashby, S. D. Hallam, *Acta Metall.* **34**, 497 (1986).
7. J. R. Marko, R. E. Thomson, *J. Geophys. Res.* **82**, 979 (1977).
8. R. Kwok, in *IUTAM Symposium on Scaling Laws in Ice Mechanics and Ice Dynamics*, J. P. Dempsey, H. H. Shen, eds., Springer (2001), p. 315.
9. R. Kwok, M. D. Coon, *J. Geophys. Res.* **111**, C11S21 (2006).
10. C. E. Renshaw, E. M. Schulson, *Nature* **412**, 897 (2001).
11. N. Golding, E. M. Schulson, C. E. Renshaw, *Acta Mater.* **60**, 3616 (2012).
12. R. Frederking, I. J. Jordaan, J. S. McCallum, in *IAHR 90: Proceedings of the 10th International Symposium on Ice, Volume 2*, Helsinki University of Technology (1990), p. 931.
13. J. Wells et al., *Cold Reg. Sci. Technol.* **65**, 314 (2011).
14. C. E. Renshaw, E. M. Schulson, *J. Geophys. Res. Solid Earth* **122**, 1088 (2017).
15. R. A. Massom et al., *Nature* **558**, 383 (2018).
16. M. H. Meylan et al., *Proc. R. Soc. A* **477**, 20210173 (2021).
17. P. Paris, F. Erdogan, *J. Basic Eng.* **85**, 528 (1963).
18. A. Engelbrektson, *Evaluation of Extreme Ice Forces on a Lighthouse in the Bothnian Bay: A Study of the Björnklack Event in April 1985*, Ice Forces Against Offshore Structures rep. no. 1, National Administration of Shipping and Navigation and University of Luleå, Sweden (1987). PT

MAKING GRADUATE ADMISSIONS IN PHYSICS MORE EQUITABLE

Physics graduate students at Michigan State University attending class (top), working on a problem set (middle), and solving an in-class problem (bottom). (Courtesy of Harley Seeley.)





Nicholas Young is a postdoc at the Center for Academic Innovation at the University of Michigan in Ann Arbor. **Kirsten Tollefson** is an associate dean in the graduate school and a professor in the department of physics and astronomy, and **Danny Caballero** is an associate professor in the department of physics and astronomy and the department of computational mathematics, science, and engineering, both at Michigan State University in East Lansing.

Nicholas T. Young, Kirsten Tollefson, and Marcos D. Caballero

Preliminary results from the revamping of Michigan State University's physics graduate admissions process suggest that the changes have made the procedure fairer for all.

Over the years, national and international physics organizations, research laboratories, and physics departments have called on leaders in the field to address issues of diversity, equity, and inclusion. In response, the physics community has built programs that increase opportunity and support for a more diverse scientific workforce and aim to address long-standing disparities in who earns degrees in physics; physics educators have developed and implemented pedagogy and curricula that provide more equitable learning opportunities for our students;[1] and physics organizations have developed and implemented codes of conduct across our organizations aimed at making meetings and conferences more inclusive.

But the issues with ensuring a diverse, equitable, and inclusive environment are systemic and pervasive. They are perpetuated often not by actions but by inactions or well-intentioned, yet misplaced, actions.[2] To make physics more diverse, equitable, and inclusive, we must address systemic issues directly, collaboratively, and reflectively. One area that we have chosen to focus on is graduate admissions. It is especially important to consider diversity there because of its potential ripple effect across science and engineering. Individuals who complete graduate physics degrees are well positioned to become scientific leaders in industry and government, and physicists who pursue academic careers will train the next generation of scientists, engineers, teachers, and even medical doctors.

The admission of graduate students to post-bachelor's physics programs is a complex and challenging system. Any graduate director, faculty member, or graduate student can recount their own vivid experience with that complicated and, quite often, opaque process.[3] At Michigan State University, we set out to understand the admissions process, determine how it was functioning, and make changes to it. We hoped to admit more diverse candidates to our program, evaluate them more equitably and holistically, and, ultimately, create a more inclusive program where each student is valued and supported in their studies. We are far from our ideal collective vision but feel that our progress toward that vision is nonetheless important to share with our colleagues.

The traditional admissions process in physics

Unlike undergraduate admissions, which is typically carried out by a centralized admissions office, individual departments or programs usually handle the review and evaluation of graduate applications. An applicant to a US physics graduate program will typically submit their CV, undergraduate transcripts, letters of recommendation, and multiple written statements covering such topics as their personal history, research experience, motivations and goals for attending graduate school, and, occasionally, how their experiences and actions contribute to fostering a diverse community.

Depending on the program, an applicant may also be required to submit their scores for the general Graduate Record Examination (GRE) and the GRE subject test in physics (GREP). Currently only around 25% of physics and astronomy graduate programs in the US and Canada require or recommend the submission of GRE scores. Most of the other programs treat score submission as optional.[4] A group of physics faculty members reviews the applications and extends offers to the selected applicants. Some departments also have postdoctoral researchers and current graduate students assist in reviewing applications, although our department does not.

From various studies conducted on the graduate admissions process in physics, we know that quantitative parts of the application, such as grade point average (GPA) and GREP scores, usually drive the admissions decision. The strength of an applicant's letters of recommendation and the specific physics courses they took as an undergraduate are also important.[5] Although that approach to admissions is somewhat successful—today the number of physics PhDs awarded annually is near the all-time high—it is inequitable. It favors applicants from groups who are already advantaged in physics and hurts applicants who are underrepresented in the field (see figure 1).

That pattern is most apparent in GREP scores, where Asian and white men tend to score higher than everyone else.[6] In the past, some physics departments required minimum GREP scores for admission,[5] which meant that applicants who were not Asian or white men were at a disadvantage.

The GREP also introduces inequities based on an applicant's financial resources. Taking the test and sending scores to each institution one applies to can cost hundreds of dollars. Moreover, applicants at smaller schools might need to travel to a testing location and potentially stay there overnight. For students working jobs at the federal minimum wage of $7.25 per hour, the cost of taking the exam can easily exceed 40 hours of take-home pay.

Further, some of our previous work found that applicants from larger departments or more selective schools tended to score higher on the GREP than applicants from smaller departments or less selective schools.[7] Because students at those larger departments or more selective schools tend to be less diverse and more affluent than the college population at large, using the GREP in admissions can further filter out many of the students whom departments are attempting to attract through diversity, equity, and inclusion initiatives.

Other sources of inequity

If the GREP were the only inequitable part of the application process, it would be easy to make admissions more equitable by removing it. Indeed, for all the above reasons, admissions committees had already begun de-emphasizing GREP scores even before pandemic disruptions, such as testing-site closures and the lack of a virtual GREP exam, which meant that an entire cohort of students who never took the test was admitted.

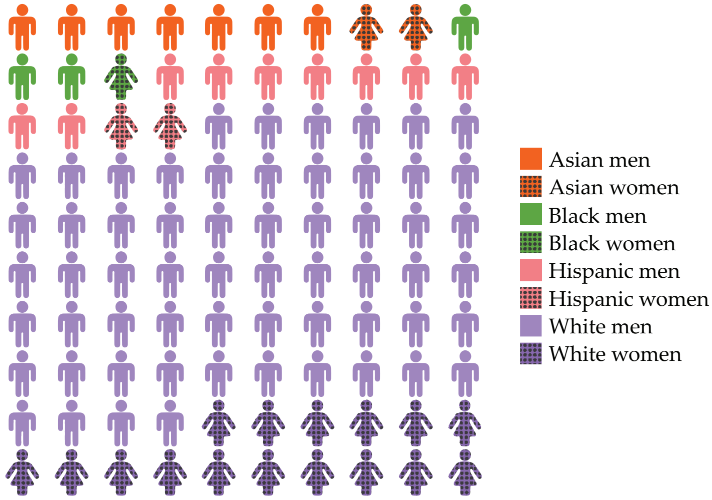


FIGURE 1. A VISUAL REPRESENTATION of the potential applicant pool for physics graduate programs. Each glyph corresponds to 1% of US graduates of various races, genders, and ethnicities who received physics bachelor's degrees from 2016 to 2020. Students who identify as American Indian, Alaska Native, Native Hawaiian, or Other Pacific Islander represent less than 1% of physics graduates and are not shown in the plot. (Data courtesy of the American Physical Society and the Integrated Postsecondary Education Data System.)

Unfortunately, that hasn't changed the admissions pattern. Even when GREP scores are removed, inequity permeates other parts of the application process.

Grades, for example, are a major determining factor in whether an applicant will be admitted, and they, too, have been found to show gender and racial differences. For example, a 2021 study conducted by researchers at the University of Pittsburgh found that students who belong to underrepresented minority groups earned lower grades than even the most disadvantaged students from any other group.[8]

In physics, various investigations have found that women earn lower grades than men. For example, one multi-institutional study found that even after accounting for prior performance, women earned lower grades than men in introductory physics.[9] That result suggests that differences in grades in those courses are more reflective of grading policies than student ability. It thus follows that selecting applicants based on GPA can hurt the admissions chances of applicants currently underrepresented in physics.

The inequity can also appear implicitly through the bias of those reviewing the applications. For example, in one 2012 study, researchers rated a male applicant for a lab manager position as more competent and hirable than a female applicant with, apart from the name, an identical application. A 2020 follow-up repeated that study with both race and gender. It again found that men were viewed as more competent and hirable than women and also revealed that white and Asian applicants were seen as more competent and hirable than Black and Hispanic candidates.[10] And since applicants are required to submit statements and letters of recommendation from faculty, the nonquantitative portions of the application are also susceptible to contributing to inequitable outcomes.

Because inequities exist throughout the entire admissions

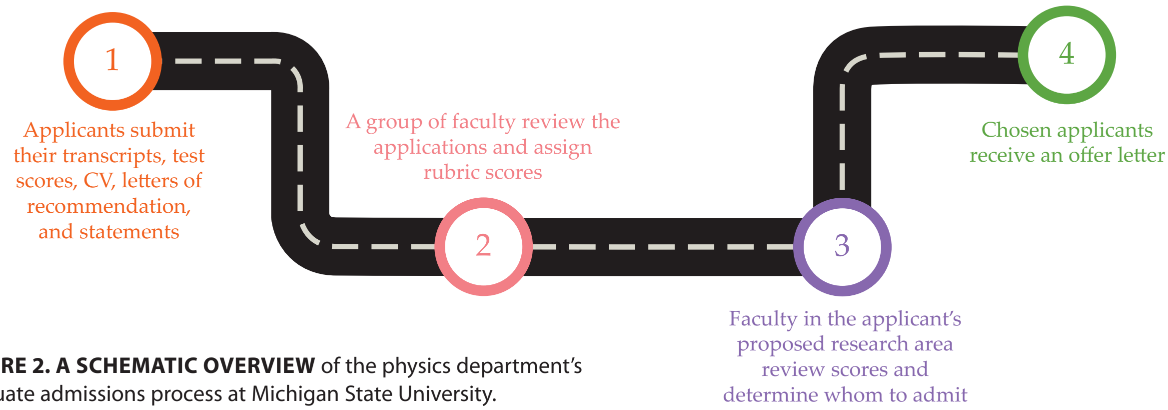


FIGURE 2. A SCHEMATIC OVERVIEW of the physics department's graduate admissions process at Michigan State University.

process, we can't simply change one part of the application process to make it more equitable. Instead, we need to consider a different approach to evaluating applicants and make the process fairer while also acknowledging that students live and learn in inequitable environments. That approach needs to consider the student as a whole and consider broadly what skills and traits an applicant needs to be successful in graduate school.

How we've changed our admissions process

In recognition of those issues, our department began to rethink its graduate admissions process in 2016. Although we expected our students to have strong math and physics skills, we also anticipated that they would be able to learn independently, take initiative, and be resilient in the face of difficulties and unexpected challenges. But our admissions process didn't have a standardized way to assess applicants on the last traits. Normally it didn't even take them into account, and if they were considered at all, they were implicitly determined from the applicant's personal and research statements.

Around the same time, physics and astronomy departments were beginning to think about how to increase diversity in their programs. Many of them started to consider the idea of assessing applicants' noncognitive traits.[11] Given the subjective nature of assessing such traits, one recommendation we received about making the process more fair to all applicants was to use a predefined rubric.[12] By using one that defines all the evaluation criteria ahead of time, applicants are compared on the same basis, and evaluators have less to debate about whether an applicant demonstrates the expected trait. Determining the evaluation criteria ahead of time can also reduce subjectivity in evaluations.[13]

The Inclusive Graduate Education Network, an NSF-funded partnership working to increase the participation of racially and ethnically marginalized students in graduate programs in the physical sciences, has conducted work on holistic admissions. After learning about their studies, our department invited two members of its management team, Casey Miller and Julie Posselt, to lead a workshop for faculty serving on our graduate recruiting committee. As a result of that workshop, faculty members decided on five categories for a rubric that aligned with both their previous experience from reviewing applications and the recommendations of the workshop leaders: academic preparation, research experience, noncognitive competencies, fit with program, and GRE scores. (Iterations of our rubric since the pandemic no longer include GRE scores.)

Each of the categories was then further divided into subcategories that mapped onto specific information about the applicant, such as their technical skills, their GPA in physics courses, and whether their research interests aligned with those of faculty members. Information to assess the subcategories, of which there are 18 in total, comes from the applicant's materials, which include transcripts, a CV, a personal statement, a research statement, and letters of recommendation. (The post-pandemic rubric, which eliminated the use of GRE scores, now contains only 16 subcategories.) To evaluate the nonacademic categories on the rubric, we asked applicants to respond to specific prompts in their personal and research statements. Those prompts broadly map onto at least one subcategory of the rubric.

One of the subcategories rates applicants on their contributions to diversity in physics through their research, teaching, or volunteering efforts (and not simply based on whether they belong to an underrepresented group in physics). Because public universities in Michigan—as in many other US states—are legally prohibited from discriminating against or granting preferential treatment to applicants based on race, sex, or ethnicity, such a scoring system ensures our admissions practices are compliant with state law.

A subset of the admissions committee then rates applicants as low, medium, or high on each subcategory, with clear criteria for what constitutes each level. By using a limited number of ratings on our rubric, we hoped to help admissions committee members avoid getting bogged down debating small differences, such as the distinction between a 3.70 and a 3.75 physics-major GPA. Although it's not included on the rubric, faculty members are also asked to note which subfields in physics and astronomy the applicant expressed interest in.

After each application has been evaluated by members of the admissions committee, a total score is calculated, based on a weighted average of the five categories. The applications, scores, and subscores are then sent to faculty representatives in each of the department's major research areas. They then make a list of applicants in their research area to whom they would like to extend an offer. Because the number of offers depends on funding and research-area needs, we do not make them based on a cutoff rubric score. Instead, we use the total score as a guide for which applicants we might want to admit.

Figure 2 presents a schematic overview of our new admissions process. Initial feedback from faculty who have served

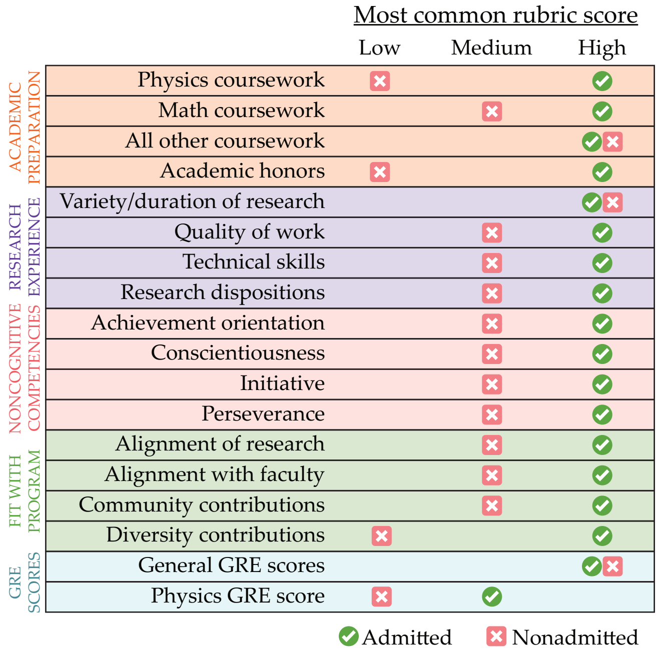


FIGURE 3. THE MOST COMMON SCORE on each rubric subcategory for admitted and nonadmitted applicants. A rating of high was the most common score for admitted applicants across the subcategories; a rating of medium was the most common score for nonadmitted applicants. Iterations of the rubric since the pandemic no longer include the two GRE-related subcategories. (Adapted from ref. 14; CC BY 4.0.)

on our admissions committee has been positive. They like that the rubric provides guidelines for how to review applications and that it defines the measures of success. They also believe that the rubric has not increased the time it takes for review, which remains between 15 and 30 minutes per application.

Did we succeed?

The goal of rethinking our admissions process was to make it more equitable. To see if that happened, we looked at the initial three years of data. The results are promising.[14] At the time of the study, the university admissions system collected only binary sex data on applicants and no racial or ethnic data. Since then, the system has been updated to allow applicants to disclose their gender, race, and pronouns if they want.

We first looked at how faculty assigned scores to the different applicants. If the rubric was useful for determining whom to admit, we would expect applicants who were admitted to have higher scores than those who were not. That was indeed what we found: Admitted applicants generally had higher scores on rubric subcategories than nonadmitted applicants. The most common rating among admitted applicants was high; among nonadmitted applicants, it was medium (see figure 3).

Next we looked at whether the rubric was equitable with respect to sex. If that were the case, we would expect males and females to have similar scores on average. Aside from a few subcategories, that is what we found, and we believe those exceptions reflect the rubric capturing known systematic issues. For example, males had higher rubric scores than females did on the GREP subcategory, and females had higher rubric scores than males on community and diversity contributions. But we've long known that males do better on the GREP than females, so it is no surprise that the rubric would measure that. Similarly, females are more likely to serve as volunteers and are often expected to take on more outreach and community-building efforts in academia. So we should not be surprised that females earned higher scores than males on those rubric sections.

We then looked at how applicants from different types of institutions performed on the rubric. Our analysis considered the overall selectivity of the institution and, as a proxy for department size, the typical number of physics degrees it awarded. Based on our experiences, we assumed that applicants from larger departments or more selective institutions had access to more resources and opportunities than applicants from smaller departments or less selective institutions. For example, an applicant from a larger department or more selective institution might have more research opportunities or have access to more advanced physics courses, and those differences might be reflected in scores on the rubric.

But aside from the GREP subcategory, we did not find any consistent differences between applicants from different types of institutions. What was most surprising was that our faculty members did not rate applicants from smaller departments lower than applicants from larger departments on the research subcategories. But prior studies have found that undergraduate students with limited research experience may not apply to graduate school in the first place,[15] which may explain why we did not find differences on those subcategories.

Lessons learned

In addition to thinking about equity in terms of rubric scores, we also considered how the rubric affected the number of female and underrepresented racial minority students who enrolled in our program. Just because applicants receive similar scores doesn't mean that admissions decisions are made along the same lines. For example, if faculty members had to choose between two comparable applicants, they might consider criteria outside the rubric to help make a distinction between the applicants. We did not find that to be the case. Since implementing the rubric, the percentage of admitted applicants who are female has more than doubled, from 13% to 31%, and the percentage of admitted applicants who are of an underrepresented racial minority group has increased from 9% to 12%. But those rates are still far from what we would hope for to achieve parity in representation.

Finally, we examined whether the rubric fundamentally changed our admissions process. We put countless hours into creating the rubric, but we still hadn't determined whether our department was basing its admissions decisions on a broader set of criteria or still relying mainly on grades and test scores. So we used machine learning to create models of our admissions process before and after we started using the rubric. Because we didn't have access to the qualitative parts of the application, such as personal and research statements for both time periods, our models used only quantitative aspects, such as the GRE scores and GPA. It also took into account the applicant's undergraduate institution and their physics subfield of interest.

We found that before we started using the rubric, our model could correctly predict whether three out of every four

See it online

Over the last 20 years, the number of students graduating with bachelor's degrees in physics has nearly doubled, but the number of positions at physics graduate programs has only grown slightly. Visit **physicstoday.org/grad-admissions** to learn about that trend and its impact on efforts to make admissions more equitable.

applicants would be admitted based on only the applicant's GREP score, GPA, and score on the quantitative portion of the standard GRE. The data from after we started using the rubric are murkier. Those three numbers are no longer determinative of whether an applicant will be admitted, which does make it seem like we are evaluating applicants on a broader set of criteria.

But even when we used the rubric scores to build a model, the resulting simulation was not able to predict whether a given student will be admitted. Perhaps the lack of a few predictive features signifies that our admissions process has become more holistic and that the rubric has created multiple routes to admission. We're currently working on determining what parts of the application are driving admissions decisions so that we can know for sure.

Using the rubric, our department has admitted more applicants from underrepresented groups in physics without increasing the time required for faculty to review applications. Based on that experience, we recommend that other physics departments implement rubrics in their admissions process to help evaluate applicants on a wider range of criteria than simply grades and test scores. But using a rubric will not result in a more equitable admissions process unless it is implemented properly. Departments need to ensure that their process also reflects a commitment to equity.

To do so, admissions committees should ensure that their members do indeed use the rubric to review each application. The committee itself should also be as diverse as possible so that it is reflective of the applicant pool.[16] Finally, we recommend that departments conduct regular reviews of their admissions processes. Just because a department has always done admissions in a certain way does not mean that they need to continue doing so, especially if the data suggest that their process is not aligned with the goals of their program.

What's next

We started using the rubric to evaluate applicants for the class of graduate students who enrolled in fall 2018. Those students are now beginning to complete our program, which means that we are just starting to understand how changes to our admissions process may have affected other areas of our program, such as time to PhD candidacy and time to completion. The initial results suggest that students admitted under the rubric are no more likely to leave the program than students admitted before we began using it.

Some might worry that reducing the role of the GREP and undergraduate GPA will lead to a weaker or less prepared incoming class and, as a result, a longer time to degree, but recent studies in engineering suggest that that hasn't happened in practice.[17] Additional work with rubric-based admissions will confirm whether those concerns are valid in physics graduate programs.

While our work suggests that rubrics with a broader range of admissions criteria make admissions more equitable, the evaluation of applicants is only one part of the process. Another part of improving equity in physics graduate education is ensuring that all groups have a fair chance in the admissions process, and doing so requires that currently underrepresented groups be included in the applicant pool. That means that future efforts at making graduate admissions more diverse and equitable need to focus on recruitment.

Moreover, once we've admitted diverse students, we need to retain them in our programs. To do so, we need to consider how our courses, qualifying and comprehensive exams, and advising and mentoring structures affect our retention efforts. At Michigan State, based on feedback from students and faculty, we've removed our qualifying exam requirement and changed the timeline for when students take their comprehensive exams. We've also added additional mentoring support for students before they form their thesis committee. We are not alone in making those types of changes: Other physics departments have also changed their exam requirements and added additional mentoring support for graduate students.[18]

The future of physics can be diverse, equitable, and inclusive if we work to make it so. Rethinking graduate admissions is one place to start.

REFERENCES

1. C. Mathis, S. Southerland, *Phys. Teach.* **60**, 260 (2022).
2. M. Dancy, A. Hodari, https://arxiv.org/abs/2210.03522.
3. J. R. Posselt, *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*, Harvard U. Press (2016).
4. J. Guillochon, "GRE requirements and admissions fees for US/Canadian astronomy and physics programs," spreadsheet available online at https://tinyurl.com/33tdfvjc (last updated 1 June 2023).
5. G. Potvin, D. Chari, T. Hodapp, *Phys. Rev. Phys. Educ. Res.* **13**, 020142 (2017).
6. C. W. Miller et al., *Sci. Adv.* **5**, eaat7550 (2019).
7. N. J. Mikkelsen, N. T. Young, M. D. Caballero, *Phys. Rev. Phys. Educ. Res.* **17**, 010109 (2021).
8. K. M. Whitcomb, S. Cwik, C. Singh, *AERA Open* (2021), doi:10.1177/23328584211059823.
9. R. L. Matz et al., *AERA Open* (2017), doi:10.1177/2332858417743754.
10. A. A. Eaton et al., *Sex Roles* **82**, 127 (2020).
11. C. W. Miller, *Status: A Report on Women in Astronomy*, January 2015, p. 1.
12. C. Miller, J. Posselt, *Physics* **13**, 199 (2020).
13. I. Bohnet, *What Works: Gender Equality by Design*, Belknap Press (2016).
14. N. T. Young et al., *Phys. Rev. Phys. Educ. Res.* **18**, 020140 (2022).
15. G. L. Cochran, T. Hodapp, E. E. A. Brown, in *2017 Physics Education Research Conference Proceedings*, L. Ding, A. Traxler, Y. Cao, eds., American Association of Physics Teachers (2017), p. 92.
16. S. F. Roberts et al., *Educ. Sci.* **11**, 270 (2021).
17. L. Stiner-Jones, W. Windl, "Work in progress: Aligning what we want with what we seek: Increasing comprehensive review in the graduate admissions process," paper presented at the 2019 ASEE Annual Conference and Exposition, 16–19 June 2019. Available at https://peer.asee.org/33592.
18. R. Barthelemy et al., *Phys. Rev. Phys. Educ. Res.* **19**, 010102 (2023). PT

STEFANO DELLA BELLA/PUBLIC DOMAIN

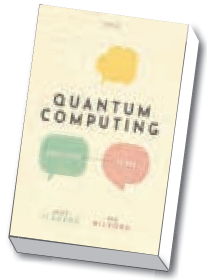
Perhaps the most famous dialog in the history of science is Galileo Galilei's *Dialogue Concerning the Two Chief World Systems*, whose frontispiece, reproduced here, depicts Aristotle, Ptolemy, and Nicolaus Copernicus (from left to right). Their astronomical systems are discussed in the dialog.

Squaring the quantum computing circle

Currently awash in funding, educational efforts, and hype, quantum information science studies how to leverage the nonclassical properties of entangled quantum states to perform certain calculations more quickly than classical computers can. Examples include factoring large numbers, which is useful for breaking encryption, and finding the ground-state energy and structure of a molecule. Authors and curriculum developers striving to help novices understand quantum information science face a daunting task, and not just because the concepts and mathematics are difficult: It's equally challenging to meet the learner where they're at.

Alice Flarend and Bob Hilborn's *Quantum Computing: From Alice to Bob* attempts to thread the needle between, in their words, "highly technical books aimed at professional scientists and engineers" and "general audience books that use almost no math." The authors assume readers are familiar with high school physics and have a mathematical background

Quantum Computing
From Alice to Bob
Alice Flarend and Bob Hilborn
Oxford U. Press, 2022. $80.00



that goes no further than trigonometry. Adding to the degree of difficulty of their deep dive, however, is their goal—which is thoughtfully implemented throughout the book—of preparing readers "to talk intelligently with the 'experts'" and "dig into the more technical aspects of the field."

The strengths and weaknesses of the book reflect the trade-offs that arise from those decisions. To facilitate communication with the quantum computing community, the authors introduce Dirac notation right from the beginning and take the time to help readers learn to translate back and forth between it and vector notation. (By contrast, many other introductions to entangled states use simplified, often pictorial, "toy" notations.) Operators are presented using matrices and several other representations, and the authors carefully connect quantum gates to those operations. To emphasize how quantum and classical computing differ, the authors introduce the classical AND and OR computing gates using the same notations they later use for their quantum counterparts. The authors work through basis changes for spin-½ systems in full mathematical detail except for some phase considerations that require complex numbers.

The authors' slow, detailed treatment of the mathematics of entangled states and algorithms, however, means that readers need to reach page 100 before they learn about any common applications of quantum information science. A snooper-detection protocol, for example, exploits the impossibility of measuring, say, an ion's x-component of spin without losing the ability to determine whether the ion's z-component was spin-up or spin-down. Similarly, it's not until page 160 that readers encounter quantum computing algorithms that enable fast searches of certain types, fast factoring of large numbers, and so on.

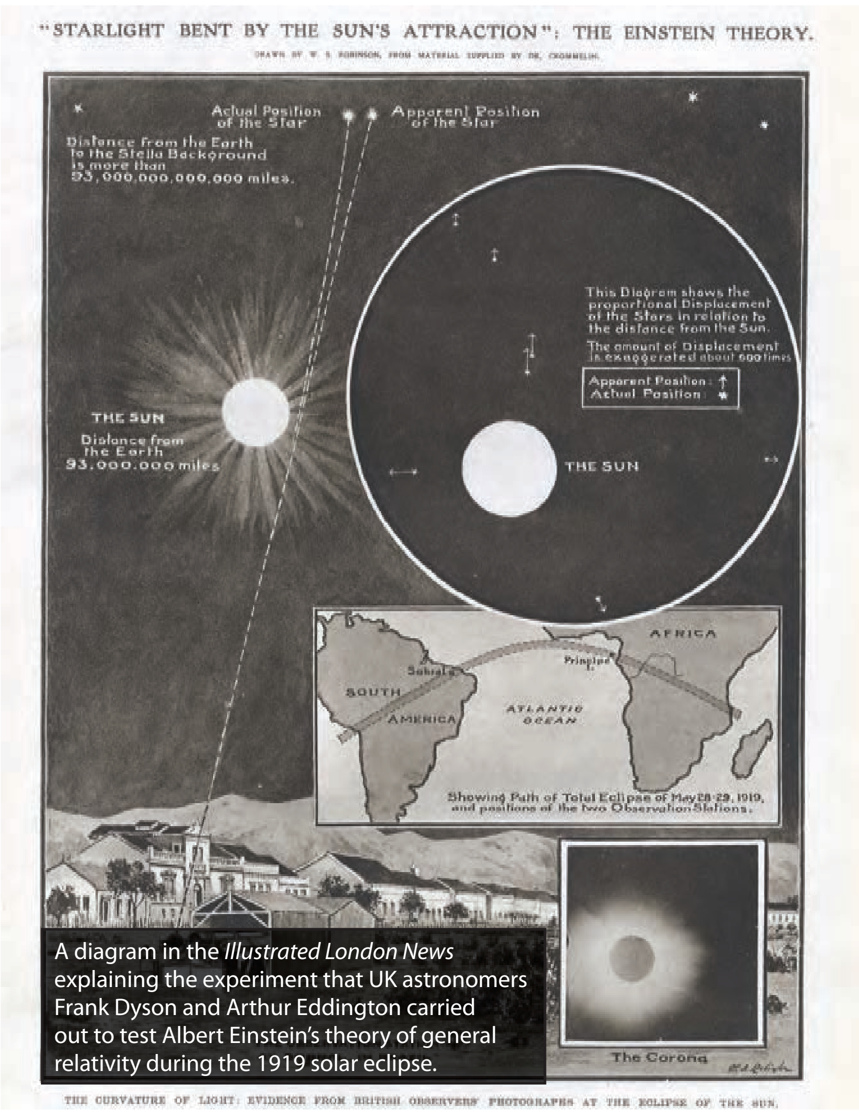
Other notable features of the book don't involve trade-offs. Aware of the limits of purely lecture-based teaching, the authors sprinkle in exercises called Try Its, which are designed to help readers work out the details of a calculation or predict what would happen in a modified version of the scenario. I found those helpful.

Less successful is their adoption of a Galileo-style dialog structure in which the authors, Alice and Bob, chat with each other and a hypothetical student named Cardy. The student interlocutor provides a way for the authors to address anticipated questions and confusions without needing to write sentences like "You might wonder why the snoop can't just reproduce the state of the qubit they measured." But Alice's and Bob's voices are so similar that they are practically indistinguishable at times, and Cardy doesn't consistently sound like a student.

Will the book reach its intended audience? It's hard for me to judge because I'm (allegedly) not a physics and linear-algebra novice. As someone who has taken undergraduate and graduate quantum mechanics courses, I found that I could skim quickly through some sections of the book but needed to work slowly through others, even before I reached the discussion of quantum algorithms. I wonder if the treatment is too dense and purely formal to hold the interest of any but the most patient and persistent reader who hasn't previously taken quantum mechanics and linear algebra courses.

A reader who hasn't previously learned about certain concepts—such as spin-½ particles, why angular-momentum-conserving interactions between them produce entangled states, and how operators rotate or reflect or otherwise redirect vectors—might find parts of the presentation dense, dryly formal, and lacking footholds in physical reality or mathematical intuition. The lack of discussion of how quantum gates are physically realized—an unavoidable decision because readers would need at least an undergraduate physics background to follow such an explanation—reinforces the dry formality. In the end, the book might be best suited for current and former physics majors who realize they need a review of spin-½ systems when learning about quantum computing. People like me.

Andrew Elby
University of Maryland
College Park



A diagram in the *Illustrated London News* explaining the experiment that UK astronomers Frank Dyson and Arthur Eddington carried out to test Albert Einstein's theory of general relativity during the 1919 solar eclipse.

W. B. ROBINSON/ILLUSTRATED LONDON NEWS/PUBLIC DOMAIN

The reality of cosmology

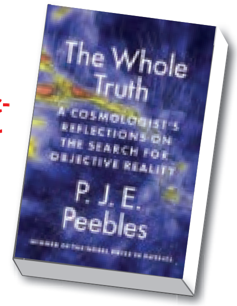
In his recent book, *The Whole Truth: A Cosmologist's Reflections on the Search for Objective Reality*, P. J. E. Peebles presents a tour de force on a rarely discussed subject: Is there objective reality in our theories of physical phenomena? Before delving into that subject in relation to cosmology, I now address the audience for the book. The inside flap of its dust jacket states that it is "essential reading for anyone interested in the practice of science." That claim is misleading. Many people who have such an interest are not well versed in physics. Reading some parts of the book will be tough sledding indeed unless one has a good background in physics.

The book starts with a fairly broad dip into the philosophical underpinnings of science, which focuses mainly on 19th- and 20th-century ideas. Along the way, Peebles uses the example provided by physical cosmology to explore what we might mean by objective reality and how we might describe it. That choice is appropriate as he made major contributions to the model now accepted by astronomers and astrophysicists as our present best approximation to objective reality in that sphere of knowledge.

As Peebles notes, physics appears to

The Whole Truth
A Cosmologist's Reflections on the Search for Objective Reality

P. J. E. Peebles
Princeton U. Press, 2022. $27.95



be independent of cultural norms. He quotes the 19th-century philosopher Charles S. Peirce, who argued that scientists converge on the same reality despite starting with different assumptions. We assume that reality operates by rules we can hope to discover. Most scientists believe—not that I have taken a poll!—that the facts are out there to be uncovered, independent of the social norms of those who choose to look. The repeatability of measurements and observations is essential to that belief. The predictive power of science is what we expect if the world operates by rules, and scientists attempt to find good approximations to those rules.

Researchers have built up over decades knowledge relevant to the formulation of our present picture of the universe. Milestones include the first recognition of the expanding universe and its predominantly hydrogen composition in the 1920s, the development of the Big Bang theory in the 1940s, the discovery of the cosmic microwave background in the 1960s, the full recognition of the problem of dark matter starting in the 1970s, and the remarkable discovery of the acceleration of the expanding universe near the turn of the 21st century.

In large part due to Peebles's contributions, those facts were incorporated into our present theory, which is usually said in shorthand to be the ΛCDM model, in which Λ denotes the cosmological constant and CDM denotes cold dark matter. The now-accelerating universe is assumed to be pushed by the cosmological constant, which was originally introduced into general relativity by Albert Einstein in the late 1910s to keep the universe static, as it was then believed to be by some astronomers. It has now been resurrected from the dustbin of history to account for the accelerating universe. Peebles does a truly admirable job of marshaling the evidence that supports the ΛCDM theory, which leaves the reader with the feeling that he is correct in concluding that the universe pretty much obeys it and that it is thus a good example of objective physical reality.

Peebles considers the heart of the theoretical underpinning to the ΛCDM to be Einstein's theory of general relativity and spends substantial space discussing the establishment of that theory. He claims that there were no precision tests of it until the 1960s. If you believe, as I do, that a test with about 1% accuracy qualifies as precise, then that last characterization is incorrect. The very first test, which Einstein applied himself immediately upon completing his theory, was to compare its prediction of the perihelion position of the planet Mercury with the extra observed advance of 43 arcseconds per century that Isaac Newton's theory of gravitation could not account for. Einstein's theory passed that test with flying colors: Its prediction for the value was within about 1% of the observed advance, a limit set by the accuracy then achievable with the relevant measurements.

The book also discusses the possible future of the ΛCDM theory. Peebles expects the search for improvements in it to more likely end in exhaustion than in major changes. He makes the point that the search for a fundamental basis for more complicated observed phenomena will continue in the future as long as society continues to support such endeavors. He further predicts that all who investigate the cosmos will arrive at the same result—an assumption, he states, that has not been challenged by any contrary evidence in at least the last century. In science, he stresses—and almost all scientists agree—we cannot prove our theories of the universe; we can only disprove them when their predictions do not agree with our measurements or observations. Peebles also mentions several times that theories in our physical-science armamentarium are incomplete. He doesn't, however, seem to discuss what would make a theory complete.

In concluding, I think that there are likely no more than a few educated people worldwide who wouldn't learn from Peebles's book. A physicist who wishes to learn the whole truth about our current knowledge of physical cosmology could accomplish that goal by reading this book and simultaneously learn a lot about related philosophy and sociology.

Irwin Shapiro
Harvard University
Cambridge, Massachusetts

NEW BOOKS & MEDIA

Silo
Graham Yost, creator
Apple TV+, 2023

The latest science-fiction series on Apple TV+, *Silo* is a detective show set in a human refuge known as the silo: a bunker with over 100 levels that protects the last 10 000 humans. If you ask to go outside, the leaders must let you go. But leaving the silo means that you die in front of the whole community. Or does it? A murder suggests there's more to that story, and an engineer named Juliette Nichols is determined to find out what it is. Like *For All Mankind*, another one of Apple's sci-fi shows, there's a dose of science in *Silo*. It's also a useful portrayal of the ways a society would have to adapt in a closed environment. —PKG

Living Histories

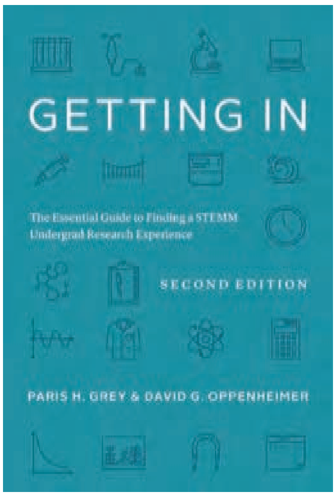
Srividya Iyer-Biswas, organizer
2020–

Conceived by Purdue University professor Srividya Iyer-Biswas, this ongoing web series aims to inspire current students by presenting brief, 15–20-minute biographical talks by established biophysicists about their own scientific journeys. Monthly presentations are streamed live and then posted on YouTube. Recent talks feature Harvard University professor Eugene Shakhnovich, who discussed his scientific upbringing as a physicist in a biological institute in the Soviet Union, and Na Ji, a professor at the University of California, Berkeley, who described how her life was shaped by her parents' experience during the Cultural Revolution in China. —RD

Getting In

The Essential Guide to Finding a STEMM Undergrad Research Experience

Paris H. Grey and David G. Oppenheimer
U. Chicago Press, 2023 (2nd ed.). $99.00



In the second edition of their user's manual for undergraduate research experiences, lab manager Paris Grey and principal investigator David Oppenheimer present an exhaustive look at the "hidden curriculum" behind lab culture. The book is divided into two parts. The first provides an overview of why research experience is valuable, describes research culture, and advises students on proper expectations for lab experiences. The second walks students through the process of searching, applying, and interviewing for research positions. *Getting In* is encyclopedic, perhaps to a fault: Although it provides a wealth of knowledge on all aspects of the undergraduate research experience, one wonders if its length and high-level language might intimidate its intended readership. —RD PT

NEW PRODUCTS

Focus on cryogenics, vacuum equipment, materials, and semiconductors

The descriptions of the new products listed in this section are based on information supplied to us by the manufacturers. Physics Today can assume no responsibility for their accuracy. For more information about a particular product, visit the website at the end of its description. Please send all new product submissions to ptpub@aip.org.

Andreas Mandelis

Thixotropic epoxy

Master Bond EP5LTE-100 is a one-part, non-premixed, frozen epoxy with a very low coefficient of thermal expansion (CTE), a high glass transition temperature (T_g), and a very high modulus. Its thixotropic paste consistency makes it suitable for bonding, sealing, and gap filling, especially in optoelectronics and applications that need high dimensional stability and good heat resistance. EP5LTE-100 passes NASA low outgassing specifications and provides a tensile modulus of above 10^6 psi at room temperature. Minimal shrinkage once cured and the low CTE of 8–12 × 10^{-6}/°C enable precise alignment for bonding dissimilar substrates with low CTEs. With a T_g between 120–125 °C and a service temperature range from −60 °C to 175 °C, EP5LTE-100 is effective in high-temperature environments. Its volume resistivity of more than 10^{14} Ω-cm makes it a reliable electrical insulator; it also resists water and damp heat. ***Master Bond Inc***, *154 Hobart St, Hackensack, NJ 07601-3922, www.masterbond.com*

Model I dilution refrigerator

Zero Point Cryogenics presents Model I, which is a complete rethinking of the traditional dilution refrigerator, designed at every stage to make operation simple and straightforward. Samples can be exchanged rapidly from the top plate, or the fridge can be opened completely when rewiring is required. With a cooldown time of less than 24 hours, rapid warm-up heaters, and reliable long-term operation, Model I keeps experiments running around the clock. The compact footprint of Model I allows users to add a dilution refrigerator where previously it had never been possible; plus the intuitive, easy-to-use touch-screen control means users do not have to be low-temperature experts to access mK temperatures. ***Zero Point Cryogenics Inc***, *9773 45 Ave NW, Edmonton, Alberta, Canada, T6E 5V8, www.zpcryo.com/model-i*

Electrical characterization cryogenic system

Lake Shore Cryotronics has launched an all-in-one system for conducting material and device research in the range of 77–500 K. The fully integrated CryoComplete cryogenic characterization system provides everything experimental physicists need to make temperature-dependent, low-level electrical measurements out of the box: a MeasureReady M81-SSM synchronous-source measure system optimized for low-level DC measurements, including a BCS-10 balanced DC/AC current source module and a VM-10 differential DC, AC, and lock-in-capable voltmeter module; an Environment by Janis VPF-100 sample-in-vacuum LN_2 cryostat with a prewired sample mount and cabling; a Model 335 temperature controller with a precision-calibrated silicon diode temperature sensor; and a computer with MeasureLINK test-control and data-charting software for automating entire experiments. The system's M81-SSM components can be used to stimulate and measure small detector photocurrents of a diode that is mounted in the cryostat and illuminated by a laser source. ***Lake Shore Cryotronics Inc***, *575 McCorkle Blvd, Westerville, OH 43082, www.lakeshore.com*

Compact mobile cryogenics

Attocube has introduced its IGLU helium compressor for Gifford–McMahon and pulse-tube cryocoolers. According to the company, the compressor system is the first in the world designed to fit into 19-inch racks. It offers a new level of compactness, ease of integration, and mobility to cryogenic systems with target temperatures below 4 K. The system's high efficiency is based on the patented IGLU technology, a new method to compress the refrigerant for cryogenic systems. In a clean, robust process, pure compression of helium takes place without the need to add oil. Thus, bulky, fragile oil separation and filtering or absorbing methods are not required. The new pumping technology saves energy: It requires only a low electrical energy input of around 1 kW and generates a low level of heat. Therefore, it can be used in most vented or climatized rooms without additional cooling infrastructure. Power and cold head speed are remotely adjustable via Ethernet. ***Attocube systems AG***, *Eglfinger Weg 2, 85540 Haar, Germany, www.attocube.com*

Intelligent interface for scroll pumps

Pfeiffer Vacuum's HiScroll series consists of three dry, hermetically sealed scroll pumps with a nominal pumping speed of 6–20 m^3/h. To enhance safety and convenience, the company now offers the AccessLink intelligent accessory interface for use with the pumps. Previously, accessories had to be configured by hand. AccessLink permits the use of a wide range of optional accessories automatically detected by the HiScroll's electronics. For example, a new gas ballast valve detects specific process requirements. The vacuum pump controls the valve according to a specific time interval or—depending on the inlet pressure—in conjunction with an additional accessory: the HiScroll integrated sensor RPT 010. New vacuum safety valves enhance operating reliability by preventing a rise in pressure at the vacuum flange when the pump is switched off. HiScroll pumps can be used in mass spectrometry, electron microscopy, and surface analysis; in accelerators and laboratory applications; and in semiconductor technology, coating, and gas recovery. ***Pfeiffer Vacuum Inc***, *24 Trafalgar Sq, Nashua, NH 03063, www.pfeiffer-vacuum.com*

Automated 3D areal surface measurements

Bruker has developed two new white-light interferometry systems for research and production. According to the company, since the stand-alone ContourX-1000 and NPFLEX-1000 optical profilometers integrate an automatic focus-finding capability, they enable fast, automated areal measurements of surface texture and roughness. The one-click Advanced Find Surface feature with autofocus and auto-illumination eliminates the complexity of manually registering the surface before each measurement. Combined with the company's self-adapting Universal Scanning Interferometry measurement mode and guided, simplified VisionXpress interface, that feature enables precise metrology on any surface at increased throughput. With user-friendly automated measurement and analysis recipes, the ContourX-1000 offers accurate, precise metrology for high-volume production facilities in semiconductor and optoelectronic processing, advanced packaging development, and medical device manufacturing. The NPFLEX-1000 combines the ease-of-use advantages of the ContourX-1000 with a large-form-factor gantry and swivel head design. It delivers a gauge-capable solution for accessing difficult orientations on large parts in quality assurance and control of precision machining in manufacturing. ***Bruker Nano Surfaces and Metrology***, *3400 E Britannia Dr, Ste 150, Tucson, AZ 85706, www.bruker.com*

High-voltage accessory for AFM

Oxford Instruments Asylum Research has presented its NanoTDDB (nanoscale time-dependent dielectric breakdown) high-voltage accessory for the Jupiter XR atomic force microscope (AFM). The technique that uses the NanoTDDB accessory measures the voltage at which a material undergoes dielectric breakdown. Users can apply constant or ramped biases up to ±220 V while monitoring current through a conductive AFM probe. According to the company, the spatial resolution of the AFM tip enables local measurements of dielectric breakdown of about 20 nm on much smaller length scales than possible with conventional probe stations. The NanoTDDB accessory can measure breakdown voltages up to ±150 V on small and large samples, such as the 200 mm wafers that the large-sample Jupiter XR can accommodate. The accessory expands the range of electrical characterization tools available on the Jupiter XR and allows for advanced measurements in the fields of semiconductors, 2D materials, thin films, and polymers. ***Oxford Instruments Asylum Research***, *7416 Hollister Ave, Santa Barbara, CA 93117, https://afm.oxinst.com*

Cold-cathode and wide-range gauges

Leybold has released a series of compact, robust, reliable cold-cathode and wide-range gauges for measuring rough to high vacuum. The active Penningvac series PTR 90 and 225 RN have been enhanced to increase their service intervals in harsh applications. The Penningvac PTR 225 RN pure-filament Pirani cold-cathode gauges use the Penning principle of cold-cathode ionization. Their measuring range of 10^{-2} to 10^{-9} enables measurements into the high-vacuum region. The Penningvac PTR 90 RN cold-cathode gauges use the inverted magnetron principle of cold-cathode ionization and a Pirani gauge. Combining the technologies into a single compact sensor increases the measuring range and enables a reliable pressure measurement from atmosphere to 1×10^{-9}. The Pirani controls the on-off function of the cold cathode automatically. That eliminates the risk of running the cold cathode at too high a pressure, which could reduce its life span. ***Leybold GmbH***, *Bonner Str 498, 50968 Cologne, Germany, www.leybold.com*

OBITUARIES

Frank Drake

Few researchers can truthfully say they pioneered an entire field of scientific research. Frank Drake, who died on 2 September 2022, at home in Aptos, California, was one. It was Drake who carried out the first astronomical search for what are now called technosignatures—signatures of technology and hence life beyond Earth. In doing so, he not only initiated the search for extraterrestrial intelligence but also shaped the field of astrobiology, the study of life of any kind in the universe.

Drake was born on 28 May 1930 on the South Side of Chicago. His father was a chemical engineer for the city and often brought home gadgets for his son that ended up in the boy's basement "lab." Frequent bike trips to the city's Museum of Science and Industry fired Drake's imagination, and a large collection of L. Frank Baum's Oz books gave him a taste for thinking about other worlds and the creatures that might inhabit them. While an undergraduate at Cornell University, Drake attended a lecture by visiting astronomer Otto Struve and was introduced to the possibility that planets and life might be common in the universe. After receiving his BA in engineering physics in 1952, Drake went on to earn his MS in 1956 and PhD in astronomy in 1958 at Harvard University under Cecilia Payne-Gaposchkin, arguably the founder of modern astrophysics.

In 1958 Drake took a position at the recently formed National Radio Astronomy Observatory in Green Bank, West Virginia. There he carried out groundbreaking studies of Venus's temperature and Jupiter's radiation belts. Drake had not, however, given up on his fascination with other planets and their potential for life. In 1960 he used the 85-foot Tatel radio dish to conduct the first experiment designed to look for life beyond Earth. Called Project Ozma—after the princess of Oz from the Baum books—Drake's effort demonstrated the feasibility for, and established the logic of, such telescopic astrobiological efforts.

Because of his work with Project Ozma, Drake was soon contacted by the National Academy of Sciences to lead a workshop on interstellar communications. In drawing together an agenda for the event, which would be attended by Nobel Prize–winning chemist Melvin Calvin and a young Carl Sagan, among others, Drake famously broke the problem into seven factors whose product yielded the number of radio-capable civilizations in the galaxy. That became the famous Drake equation, which remains one of the most well-known scientific formulas in the popular imagination. It became a backbone for the fields of astrobiology and technosignatures by establishing the key intermediate subproblems associated with finding life of any kind. Those items included detecting "exoplanets" and then finding them in the so-called habitable zone, where liquid water can exist on a world's surface. Drake lived to see both terms find observationally validated values.

While Drake initiated the field of searching for intelligence among the stars, he also created the concept of sending information about human civilization to potential civilizations. In 1974 he used the Arecibo radio telescope, the most powerful in the world, to send a signal to the globular cluster Messier 13, which is 6.8 kiloparsecs away. The signal contained a rasterized message providing, among other things, the atomic numbers of elements involved in DNA, the size of an average human male, and a graphic of the solar system. Drake was also instrumental in the design of the Pioneer plaques carried on *Pioneer 10* and *Pioneer 11* and the Golden Record affixed to the two Voyager spacecraft. Those artifacts have carried information about Earth and humanity beyond the solar system.

After his time in Green Bank, Drake served briefly as the chief of the lunar and planetary sciences section at the Jet Propulsion Laboratory. In 1964 he took a faculty position at Cornell University. He was the head of the Arecibo Observatory from 1966 to 1968 and director of Cornell's National Astronomy and Ionosphere Center, which managed Arecibo's operation, from 1970 to 1981.

In 1984 Drake moved to the West Coast to become dean of the division of natural sciences at the University of California, Santa Cruz. That same year he also joined the SETI Institute, and he became director of its Carl Sagan Center for the Study of Life in the Universe in 2004.

Drake's scientific accomplishments were well recognized by colleagues. He

Frank Drake

SETH SHOSTAK/SETI INSTITUTE

was vice president of the American Association for the Advancement of Science in 1973, president of the Astronomical Society of the Pacific from 1988 to 1990, and chair of the National Research Council's Board on Physics and Astronomy from 1989 to 1992.

Beyond his soft-spoken nature, his kindness, and his good humor, Drake will be remembered for his intellectual bravery. He championed SETI at a time when it was often considered marginal and prone to the "giggle factor" associated with UFOs and little green men. In 2020 the astronomical community's decadal survey made as its top priority a telescope designed to hunt for biosignatures. Given that technosignatures are just one specific kind of biosignature, the world is finally catching up with Frank Drake and his daring Project Ozma.

Adam Frank
University of Rochester
Rochester, New York
Jill Tarter
SETI Institute
Mountain View, California
Jason Wright
Pennsylvania State University
University Park PT

Pierre Gembaczka is a lead data scientist at Krohne in Duisburg, Germany. **Lukas Krupp** is a doctoral student in machine learning for embedded systems at the Fraunhofer Institute for Microelectronic Circuits and Systems in Duisburg.

The internet is full of things

Pierre Gembaczka and Lukas Krupp

The devices, or "things," that communicate and share data via the internet are part of a network that's becoming increasingly connected.

It's often the little things that make life easier. Think wearables, such as fitness trackers and smartwatches, or smart-home devices, such as lighting control systems, thermostats, robotic vacuum cleaners, and, of course, the voice-controlled, intelligent personal assistants that have already made their way into many living rooms. People may benefit from those and countless other devices every day, but few know exactly what's behind them and how they work. This Quick Study pulls back the covers.

The internet of things (IoT) connects all small, intelligent systems that are already an integral part of many sectors in the global economy—from industry to agriculture. The IoT is the name of the network of physical objects equipped with sensors, software, or other technology that empowers them to exchange information with other objects. Such devices typically use microcontrollers for data processing to minimize cost and energy consumption. Most of them can even operate on battery power.

The IoT has its origins in the work of Kevin Ashton, who coined the term in 1999 during a presentation he gave as an assistant brand manager at Procter & Gamble. The company's brown lipstick always seemed sold out to him even though it was still frequently available in stock. In the presentation, Ashton brought up the idea of placing RF identification tags on every product in the company in order to identify and track it through the supply chain. The IoT was conceived.

Today, it has become an enormous market, split into numerous specializations. Recent forecasts expect the IoT to grow from $478.36 billion in 2022 to $2465.26 billion in 2029. That's on par with the value of today's automotive-manufacturing industry worldwide.

Communication is the key

The IoT's things are any devices that can be assigned an internet-protocol address and can transfer data wirelessly over a network. For direct communication between objects over short distances, consumer electronics technologies, such as Bluetooth and ZigBee, are common, especially in applications where data volume is low and little energy is consumed. For local-area networking and connecting multiple devices to the internet, Wi-Fi is a prominent example. Other technologies included in modern smartphones, such as near-field communication for short distances and cellular communication for global networking, are also common in IoT devices.

Automated information exchange between end devices, such as machines or central control units, is referred to as machine-to-machine communication. An example is the communication between robots completing a common task, such as assembling products. But it's also possible to integrate a remote bank of computers, known as a cloud, to manage information from distributed devices in a central location. Several cloud service providers are now available, including Microsoft's Azure IoT Hub and Amazon's AWS IoT services.

Such IoT technology is perhaps most interesting to the industrial sector of the economy, in which the term industrial internet of things (IIoT) is now common parlance. In contrast to consumer-oriented devices, IIoT ones focus on industrially relevant applications. An example is the control of industrial robots in assembly lines awaiting sensor data to be sent to a controller before they can perform their task. Built into such control is a so-called deterministic latency, which allows a robot to react to its environment on time. Apart from such real-time constraints, many industrial applications operate in harsh conditions. Unlike a smart-home environment, in which the only information exchange may involve a few light switches, a smart factory, like the one shown here, may have dozens of machine parameters and sensor signals that must be constantly monitored and frequently processed.

Driven by the increasing distribution of IIoT devices throughout production plants, predictive maintenance and condition monitoring are the current trends in industrial fields. Thanks to that monitoring, companies can predict machine failures and defects in manufactured products at early stages—even before parts actually break. That's an essential task in a smart factory, and it's key to resource planning and process control. The multinational European aerospace company Airbus, for instance, has launched a digital-manufacturing initiative known as Factory of the Future as a way to increase its production capacity. To reduce errors and improve workplace safety, sensors are currently integrated into manufacturing tools and machines, and employees use wearable technologies, such as ecom's smart eyeglasses Visor-Ex 01.

Machine learning

Machine learning (ML) algorithms, such as artificial neural networks, play an important part in the IoT. The huge number of sensors distributed in countless devices around the world and the increasing extent of networking between them generate a tremendous amount of data. Fortunately, ML methods are capable of analyzing that data to extract information and infer what's going on. Voice-controlled assistance systems, for in-

MACHINE-TO-MACHINE COMMUNICATION. Robots coordinate when to fetch or screw together their components in this automobile assembly plant. They can autonomously schedule and optimize the assembling processes to minimize energy consumption and production time. Likewise, they can analyze their own condition using sensors embedded inside them and predict when they need maintenance or repair. (Courtesy of BMW Werk Leipzig/CC BY-SA 2.0 DE.)

stance, use ML to process and understand human speech. Likewise, sound-detection systems analyze environmental sounds, such as traffic on a highway, to interpret any detected patterns. (For ML applications to weather forecasting, see Physics Today, May 2019, page 32.)

Two main approaches exist for designing a system architecture that integrates ML and the IoT. The first approach collects data in a central location such as a cloud server. Ordinarily, that data must first be transferred from the actual device before it can be processed. Unfortunately, such an enormous amount of data streaming from IoT devices to the cloud server consumes a lot of energy. What's more, any interruption to the connection may also cause the system to malfunction. To counter those potential problems, engineers have started migrating ML methods directly into IoT devices. The idea behind that second approach, which is now gaining increasing popularity, is to perform as much of the data processing as possible inside the particular IoT device at issue. And it dramatically reduces the amount of transmitted data.

The new generation of devices embodies a paradigm often described as the artificial intelligence of things (AIoT), which integrates the ML algorithms directly inside physical objects and products. The ML methods and algorithms that produce the integration are grouped under the terms TinyML or Edge AI. Devices that benefit from the local intelligence include surveillance cameras that do not require the transmission of raw images to a cloud server and autonomous robots in hostile environments that can detect and respond to threats all on their own.

High-performance hardware

Processing high volumes of data using ML requires enough computing power for the system to extract information and make a decision in a feasible amount of time. Specialized computing hardware, such as graphics processing units, can accelerate the data processing. Faster processing speed leads to lower energy requirements and ultimately an increased amount of data that can be processed. That capability supports outsourcing complex applications without the need for a permanent connection to a cloud server.

New ways for robots to predict failures have emerged. Multiple high-speed data streams from cameras, microphones, and other sensors must be processed in parallel, such that the robots can understand their environment—by seeing through a camera, feeling through a pressure sensor, or hearing through a microphone.

What could the next generation of AIoT devices look like? One possibility is self-learning devices, such as wearables that can adapt to a specific user or to an application. For large tasks, such as self-driving cars, several devices could even work and learn together (see the Quick Study by Colin McCormick, Physics Today, July 2019, page 66). Bringing this vision to reality will require an ML framework that can run on arbitrary AIoT devices, handle training, and enable the integration of hardware-acceleration mechanisms for ML. Researchers at the Fraunhofer Institute for Microelectronic Circuits and Systems in Germany—where one of us (Krupp) studies—have already implemented that framework. Artificial Intelligence for Embedded Systems (AIfES) is open source and developed in the C programming language so that it can run on any hardware.

For a sense of what a self-learning device could look like, consider Fraunhofer's smart power-sensor demonstrator (see www.aifes.ai), which is used to monitor the condition of a machine. In many applications, the machine's current power consumption provides detailed information about its operation, energy efficiency, and maintenance requirements due to wear or damage. The sensor can be attached directly to the machine, and the different operating states can be trained. The sensor is configured via a smartphone or tablet connected by Bluetooth. So configured, the sensor does not have to constantly send data—and does so only when changes occur in the operating state. That saves energy, and there's little risk of communication failure.

Additional resources

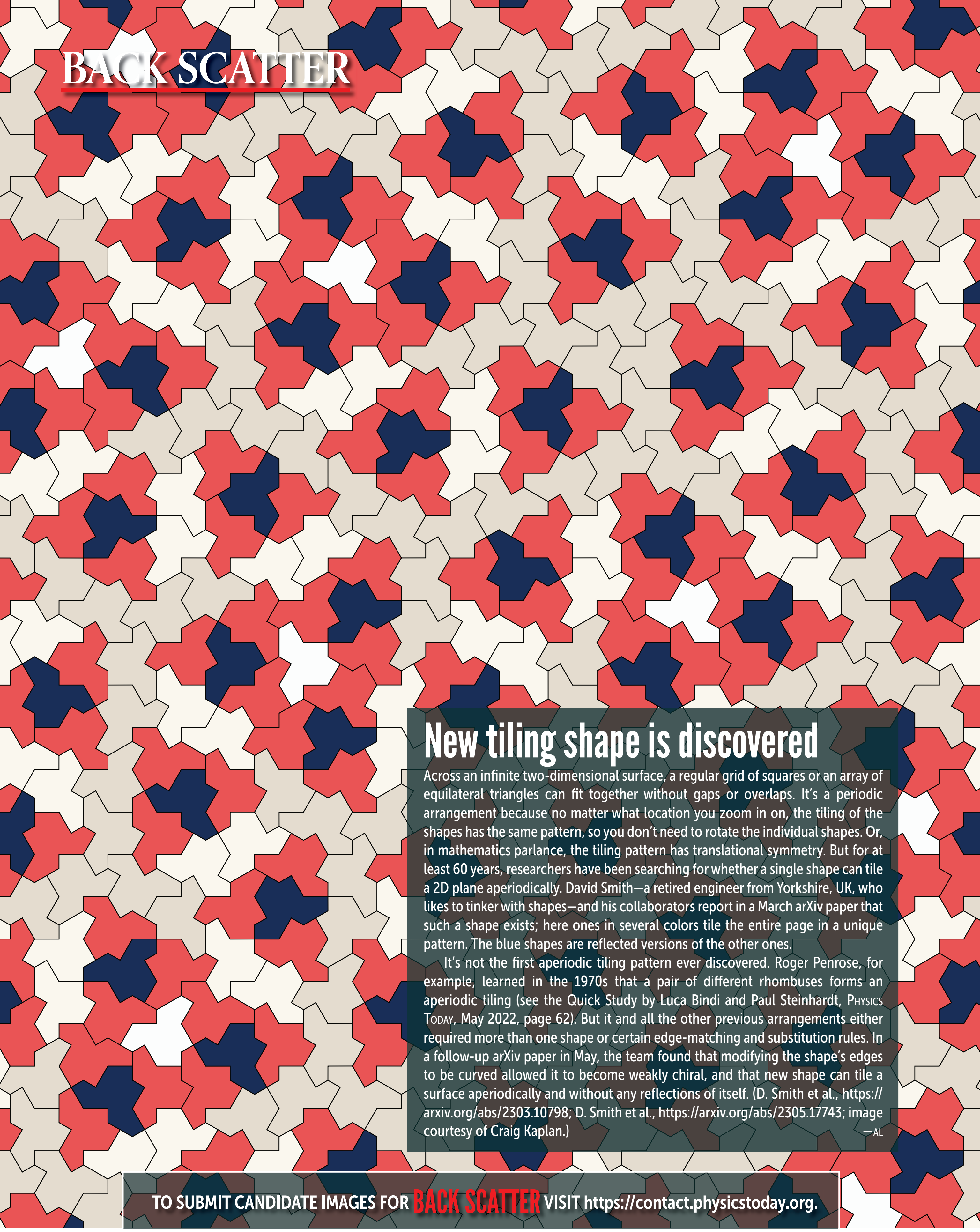
► Fortune Business Insights, *Internet of Things (IoT) Market Size, Share & Covid-19 Impact Analysis, . . . , 2023–2030* summary (2023), www.fortunebusinessinsights.com/industry-reports/internet-of-things-iot-market-100307.

► N. Dew, "Lipsticks and razorblades: How the Auto ID Center used pre-commitments to build the Internet of Things," PhD thesis, Naval Postgraduate School (2003).

► P. Jama, "The future of machine learning hardware," *Hacker-Noon* (3 September 2016).

► Fraunhofer Institute for Microelectronic Circuits and Systems, "Wireless Current Sensor for Condition Monitoring." PT

BACK SCATTER



New tiling shape is discovered

Across an infinite two-dimensional surface, a regular grid of squares or an array of equilateral triangles can fit together without gaps or overlaps. It's a periodic arrangement because no matter what location you zoom in on, the tiling of the shapes has the same pattern, so you don't need to rotate the individual shapes. Or, in mathematics parlance, the tiling pattern has translational symmetry. But for at least 60 years, researchers have been searching for whether a single shape can tile a 2D plane aperiodically. David Smith—a retired engineer from Yorkshire, UK, who likes to tinker with shapes—and his collaborators report in a March arXiv paper that such a shape exists; here ones in several colors tile the entire page in a unique pattern. The blue shapes are reflected versions of the other ones.

It's not the first aperiodic tiling pattern ever discovered. Roger Penrose, for example, learned in the 1970s that a pair of different rhombuses forms an aperiodic tiling (see the Quick Study by Luca Bindi and Paul Steinhardt, Physics Today, May 2022, page 62). But it and all the other previous arrangements either required more than one shape or certain edge-matching and substitution rules. In a follow-up arXiv paper in May, the team found that modifying the shape's edges to be curved allowed it to become weakly chiral, and that new shape can tile a surface aperiodically and without any reflections of itself. (D. Smith et al., https://arxiv.org/abs/2303.10798; D. Smith et al., https://arxiv.org/abs/2305.17743; image courtesy of Craig Kaplan.)

—AL

TO SUBMIT CANDIDATE IMAGES FOR BACK SCATTER VISIT https://contact.physicstoday.org.