PHYSICS TODAY

August 2021 • volume 74, number 8

A publication of the American Institute of Physics

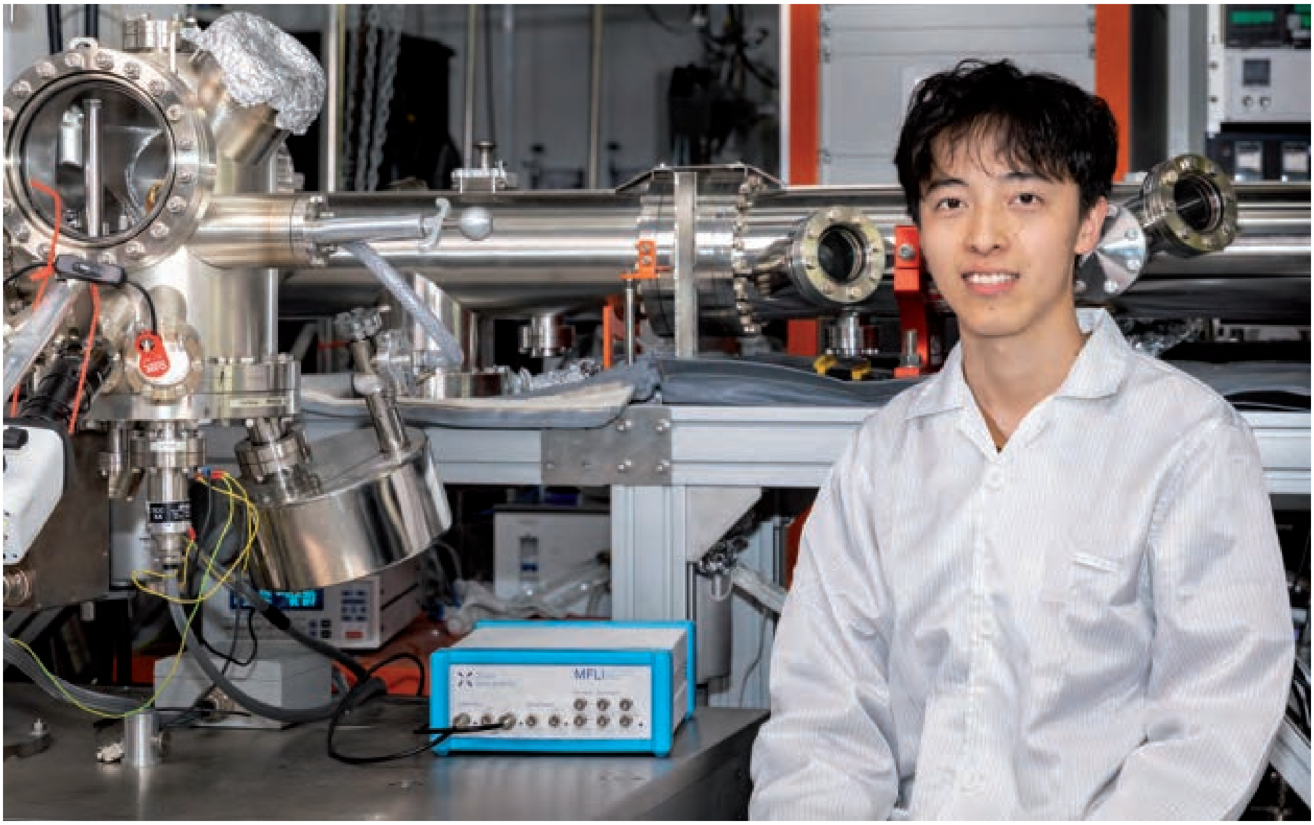
Planetary nurseries

Europe's bet on graphene

Twisting viscoelastic fluids

Chip-sized particle accelerators


Monitor Deposition in Real Time
Ellipsometry offers solutions that improve your process with live feedback.
+OPTIMIZE DEPOSITION PROCESS
Determine thickness and optical properties of films and multi-layer stacks for metals and dielectrics
+IMPROVE QUALITY CONTROL
Detect variations as they occur with live feedback during the deposition process
+MONITOR GROWTH KINETICS
Sub-angstrom thickness sensitivity provides additional information about film nucleation, surface conditions and other process conditions
J.A. Woollam



Yumeng Yang, Spin InfoTech Lab, ShanghaiTech University

Faster and more reliable

We congratulate the Spin InfoTech Lab led by Prof. Yumeng Yang at ShanghaiTech University for realizing fast and energy-efficient magnetization switching. Through charge-spin conversion based on the anomalous Hall effect in ferromagnetic materials, this result can be used in a magnetic random access memory (MRAM) bit to improve reliability, memory density, and speed. This work will greatly benefit the commercialization of next-generation MRAM technology, which is one of the most promising candidates for high-performance neuromorphic computing.

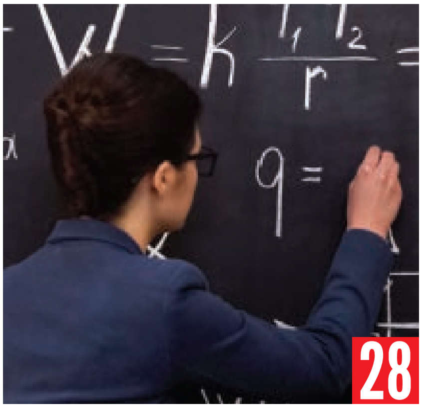
We are excited to continue our collaboration and look forward to further spectacular results.

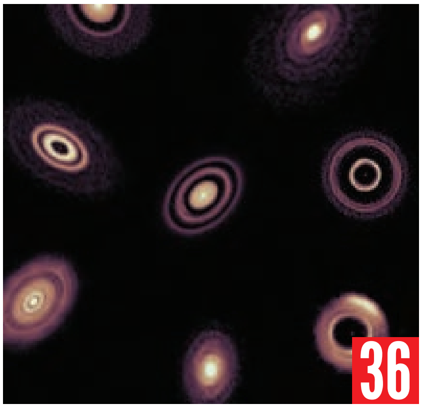
Zurich Instruments

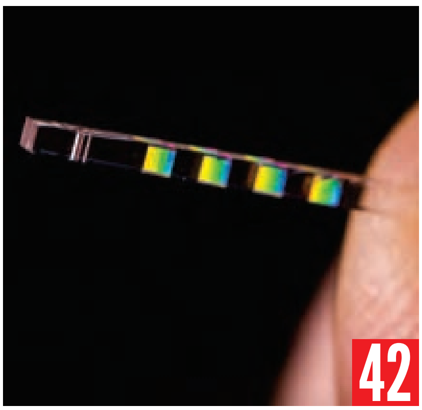
PHYSICS TODAY

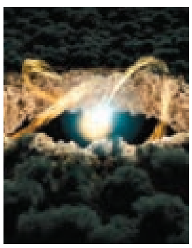
August 2021 | volume 74 number 8

FEATURES





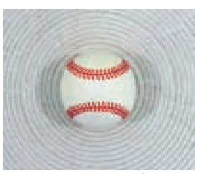




ON THE COVER: This artist's impression depicts a protoplanetary disk surrounding a star. Material from the disk flows along the star's magnetic field lines to strike the star's surface. The ensuing flares can be observed both directly and as echoes once the light reaches and illuminates the inner disk. Timing the echoes yields the radius of the inner disk. Turn to the article by Sean Andrews on **page 36** to learn how new observations are enriching our understanding of the disks of gas, dust, and ice that nurture planets. (Image courtesy of NASA/JPL-Caltech.)

Recently on PHYSICS TODAY ONLINE

www.physicstoday.org

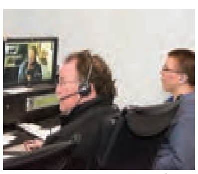


DEAN HOCHMAN/CC BY 2.0

Physics of baseball

Major League Baseball recently started cracking down on pitchers' alleged use of tacky materials to give the ball extra spin. But how much of an advantage does an enhanced grip confer? Physics Today's Andrew Grant investigates what we know about the relationship between pitch mechanics and ball trajectory.

physicstoday.org/Aug2021a



LAUREN HUGHES/NASA

Deaf scientists

Hard of hearing and deaf people are still under-represented in STEM fields, but the availability of technology and interpreters has lowered traditional barriers. Deaf scientists at NASA and CERN discuss their experiences getting access to necessary services and the needs of the deaf community in STEM.

physicstoday.org/Aug2021b

YANCOASTERMAN/CC BY-SA 3.0

Loop-the-loop

Despite its status as a classic demonstration of energy conservation, balls that travel around a loop-the-loop track lose energy. That loss is usually dismissed as arising from nonconservative forces. Physics Today's Madison Brewer reports on a recent study of the motion's complicated physics.

physicstoday.org/Aug2021c

Physics Today (ISSN 0031-9228, coden PHTOAD) volume 74, number 8. Published monthly by the American Institute of Physics, 1305 Walt Whitman Rd, Suite 300, 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 300, 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

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

Editor-in-chief
Charles Day *cday@aip.org*

Managing editor
Richard J. Fitzgerald *rjf@aip.org*

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

Editors
Ryan Dahn *rdahn@aip.org*
Toni Feder *tf@aip.org*
Heather M. Hill *hhill@aip.org*
Abby Hunt *ahunt@aip.org*
David Kramer *dk@aip.org*
Alex Lopatka *alopatka@aip.org*
Christine Middleton *cmiddleton@aip.org*
Johanna L. Miller *jlm@aip.org*
Gayle G. Parraway *ggp@aip.org*
R. Mark Wilson *rmw@aip.org*

Online
Paul K. Guinnessy, director *pkg@aip.org*
Andrew Grant, editor *agrant@aip.org*
Angela Dombroski *atd@aip.org*
Greg Stasiewicz *gls@aip.org*

Assistant editor
Cynthia B. Cummings

Editorial assistant
Tonya Gary

Contributing editors
Madison Brewer
Andreas Mandelis

Advertising and marketing
Christina Unger Ramos *cunger@aip.org*
Unique Carter
Krystal Dell
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 | American Institute of Physics

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

Other member organizations
Sigma Pi Sigma Physics Honor Society
Society of Physics Students
Corporate Associates

Symposia and webinars

Charles Day

This past January I attended a virtual symposium hosted by the Rice Center for Quantum Materials (RCQM). The topic was iron-based superconductivity, whose discovery in 2008 ignited an explosion of experimental and theoretical activity (see PHYSICS TODAY, May 2008, page 11). The symposium focused on four hot research areas: electron correlations and orbital selectivity, quantum criticality, nematicity, and electronic topology.

The 17 invited experts gave their talks by sharing PowerPoint slides from their laptops or desktops via Zoom. If the symposium had been held at Rice University in real life, the talks would have been pretty much the same. Indeed, my first impression of the virtual symposium was how similar it was to the live ones I attended in the Before Times.

But scrutinizing the agenda revealed some positive differences. The speakers came from seven different countries: Canada, China, Germany, Japan, the Netherlands, the UK, and the US. Such an intercontinental cast is not unusual for a big, week-long conference, such as the March meeting of the American Physical Society. For a three-day symposium without a deep-pocketed benefactor, it's rare. What's more, the four speakers who are based in China (and any attendees who are based in a country that does not fall under the US's Visa Waiver Program) were spared the time-consuming, costly, and uncertain chore of applying for a US visa in person at a US consulate or embassy.

Another positive difference had to do with the scheduling. RCQM divided the symposium into two daily sessions: a morning one from 10:00am to 11:30am US CST (convenient for most people in Europe and the Americas) and a night one from 8:30pm to 10:00pm (convenient for most people in South and East Asia). If all the attendees were in person, the wide gap between sessions would be unnecessary. But in a likely future of hybrid meeting, such a gap is thoughtfully accommodating.

Some of the properties of the RCQM symposium are shared by PHYSICS TODAY's webinar series, which launched in August 2019 with "Vacuum Chamber Modeling for Accelerators" from science software company COMSOL. The first webinars aimed to help companies tell potential customers about their products. Then, in December 2020, nine months into the COVID-19 pandemic, PHYSICS TODAY's director of sales and marketing, Christina Unger Ramos, came up with the idea to host an additional type of webinars: webinars by physicists for physicists.

The first editorial webinar aired this past February. The speaker, Marc Miskin of the University of Pennsylvania, expounded the micron-scale robots that he and his team make. If the topic sounds familiar, it is. Miskin wrote a Quick Study for PHYSICS TODAY's December 2020 issue. In the 40 minutes of his presentation, he could show far more graphic content—figures, movies, animations—than he could in his original two printed pages. Of the 715 people who registered, 324 watched it live.

My role in the webinars is to introduce the speaker, run through some housekeeping instructions, and moderate the questions that attendees submit in writing via the webinar platform, which is provided by ON24, a San Francisco–based webcasting and virtual-event company.

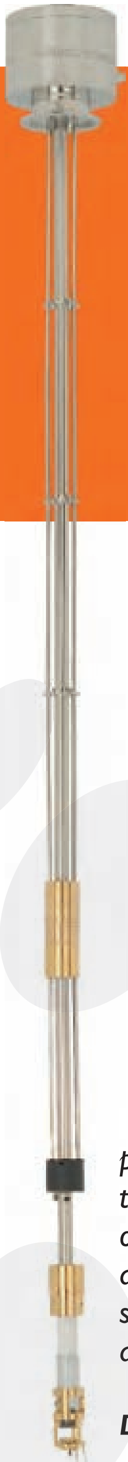
The questions I field are consistently interesting. Northeastern University's Gregory Fiete gave a webinar on 1 July on the physics and applications of rare-earth elements. Attendees asked technical questions about oxidation states and chemical analogs; they also asked about recycling old devices that contain rare earths and about the prospects for extracting rare earths from the seafloor.

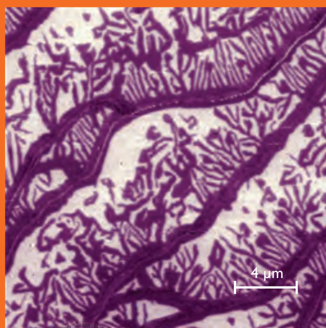
What are the implications of the success of RCQM's symposium and PHYSICS TODAY's webinars for the future of meetings? When I ask physicists about what they miss about in-person conferences, they always mention seeing their colleagues. Until I set out to write this editorial, I didn't realize the literal import of their answers. It wasn't necessarily the sessions, plenary talks, posters, and so on they missed. Rather, they wanted to talk physics again in person.

Thanks to virtual technologies, one aspect of traditional conferences—the transmission of results through talks at sessions—could readily go online. And if it does, there's no reason for physicists to convene at big downtown conference centers. What might happen is a return to the modus vivendi that Alexei Kojevnikov describes in his recent book, *The Copenhagen Network: The Birth of Quantum Mechanics from a Postdoctoral Perspective* (2020). In Europe in the 1920s and '30s, physicists visited each other's institutions. PT

Excellence in Low Temperature Imaging

LT - Scanning Probe Microscope System





KPFM image of $CaFe_2As_2$

Imaging Modes

SHPM, STM, AFM, MFM, EFM
SNOM, Conductive AFM, KPFM,
N- V Center & Confocal Microscopy

Temperature Range

10 mK - Room temperature

Essentially five reasons make researchers adapt their experimental setups to NanoMagnetics Instruments low-temperature system compatibility.

- ☑ Reduced thermal drift
- ☑ Lower noise levels
- ☑ Enhanced stability of tip and sample
- ☑ Reduction in piezo hysteresis/creep
- ☑ Probably the most obvious, the fact that many physical effects are restricted to low temperature

Can be customised to fit in any cryostat

"The LT-AFM/MFM system allows us to perform studies on functional materials to investigate magnetic, piezoelectric and morphological characteristics with nanoscale spatial resolution. The versatility of the system to switch between different measuring modes, and the possibility of working under applied magnetic fields, offers us the possibility to stablish structure-property relationships, fundamental to the understanding, design and use of materials. We are currently applying this technique to the study of vortices dynamics in layered superconductors, and the investigation of ferroelectric/ferromagnetic heterojunctions for spintronic applications."

Dr. Carmen Munuera, 2D Foundry, Material Science Institute of Madrid (ICMM-CSIC)

NANOMAGNETICS INSTRUMENTS

/NMInstruments

+44 7906 159 508
sales@nanomagnetics-inst.com
Suite 290, 266 Banbury Road Oxford OX2 7DL, United Kingdom

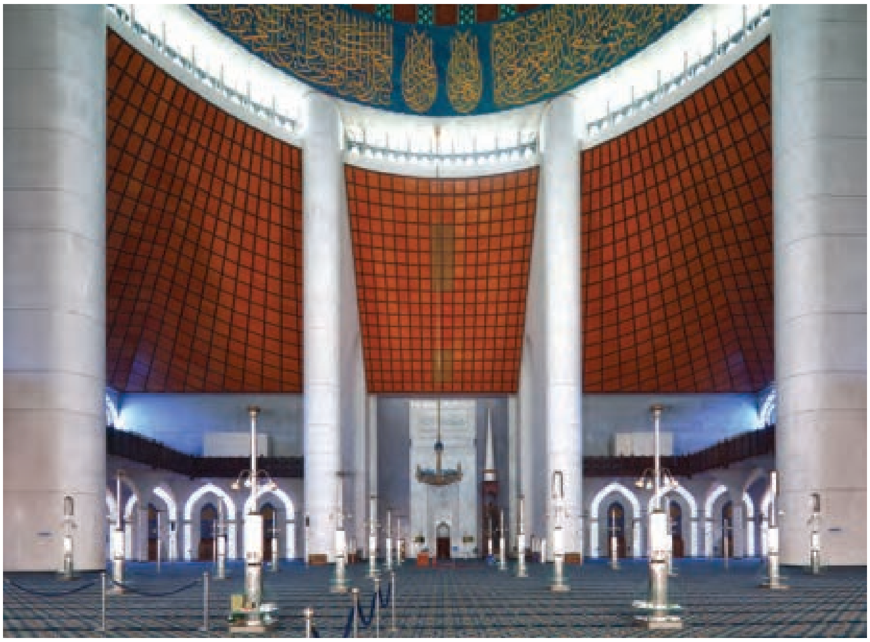
Acoustics of multiuse spaces

The article "Exploring cultural heritage through acoustic digital reconstructions" by Brian Katz, Damian Murphy, and Angelo Farina (PHYSICS TODAY, December 2020, page 32) gives an excellent review of the state of the art in computer simulation of room acoustics. However, the authors seem unfamiliar with the revolution in concert-hall, theater, and worship-space design that started with Peter Parkin and J. H. Taylor's work in London's St Paul's Cathedral[1] in 1952. Beginning with that pioneering effort, room-acoustics designers have found that the ratio of early-arriving sound energy to reverberant sound energy at the listener's ears is at least as important as reverberation for speech and music acoustics.[2,3] The usual division between early and reverberant sound is 50 ms for speech and 80 ms for music. "Usual" is an important word.

At one extreme, an acoustically "dry"—that is, nonreverberant—thrust-stage or in-the-round theater needs special electroacoustic or sound-reflecting surfaces to ensure enough early sound energy reaches a listener who is seeing the back of someone speaking rather than his or her face. Otherwise, with no real departure from previous practice, echoes usually reduce intelligibility.[4]

A more relevant and frequent application is the reconciliation of speech and music in the same space. A multiuse auditorium can, of course, have variable acoustics, but moving drapes, curtains, and sound-absorbing panels in the middle of worship services is hardly practical. Thus the greatest value of the early-to-reverberant energy-ratio concept has been for houses of worship, particularly large Christian churches and cathedrals, and for places where sound-absorbing treatments are not suited architecturally.[4]

Perhaps the best demonstration of the concept is the Sultan Salahuddin Abdul Aziz Shah Mosque, known as the Blue Mosque, in Shah Alam, Malaysia. It has the largest dome—more than 51 m in diameter, rising 75 m above the prayer hall—of any religious building. It also has the largest and most complex sound-reinforcement system. Designed by Larry Philbrick, the system uses the directional properties of loudspeakers rather than sound-absorbing treatments to control echo and reverberation and to increase clarity. Two large, central loudspeaker clusters—one for close worshippers, one for middle and far worshippers—cover the main prayer area under the dome. Additional line-source loudspeakers are built into the interior and courtyard columns. The mosque can accommodate 24 000 worshippers at one time.



B.O'KANE/ALAMY STOCK PHOTO

THE SULTAN SALAHUDDIN ABDUL AZIZ SHAH MOSQUE in Shah Alam, Malaysia, has the largest dome of any religious building. Behind the two sets of tan tiles in the central section of the sloped ceiling seen here are large clusters of loudspeakers that together cover the main prayer area under the dome.

References

1. P. H. Parkin, J. H. Taylor, *Wireless World*, February 1952, p. 54, and March 1952, p. 109, reprinted in *J. Audio Eng. Soc.* **54**, 67 (2006).
2. See, for example, M. Long, *Architectural Acoustics*, 2nd ed., Academic Press (2013), p. 748; M. D. Egan, *Architectural Acoustics*, McGraw-Hill (1988), p. 98.
3. D. L. Klepper, "An autobiographical assessment of the importance of the early-to-reverberant sound-energy ratio, 'clarity,' in speech acoustics," Pro Audio Encyclopedia (1 September 2016), www.proaudioencyclopedia.com/?p=363; L. G. Marshall, *J. Acoust. Soc. Amer.* **96**, 2251 (1994).
4. See, for example, D. L. Klepper, *J. Audio Eng. Soc.* **43**, 599 (1995), reprinted in *An Anthology of Articles on Sound Reinforcement . . .*, Audio Engineering Society (1996), p. 392; Clayton Acoustics Group, "St. Paul's Episcopal Church, Rock Creek Parish, Washington, DC" (2005), www.claytonacoustics.com/pages/portfolio/houses%20of%20worship/project%20profiles/St.%20Paul's.html.

David L. Klepper
(daveklepper@yahoo.com)
Yeshivat Beit Orot
Jerusalem, Israel

▶ **Katz, Murphy, and Farina reply:** Please note that our article was an overview of research in virtual heritage acoustics, not a presentation of modern acoustical design methods. The letter by David Klepper focuses on electroacoustic solutions to difficult acoustic situations.

Contrary to assumptions by the letter writer, we are very familiar with energy-ratio metrics in room acoustics,[1] including their limitations.[2] Although not relevant for our introductory article, we have used energy ratios—together with other measures—not only for characterizing acoustic spaces but also for calibrating and validating acoustic simulations and auralizations.[3] In particular, characterizations of Venice's La Fenice theater[4] and Paris's Notre Dame Cathedral[5] before each burned (in 1996 and 2019, respectively) inspired the name of the Past Has Ears (PHE, for the constellation Phoenix) project.

These days, other measurable quantities are often preferable to energy-ratio metrics and more reliable as refined design criteria, especially regarding natural room acoustics. For example, temporal and spatial energy-repartition measures, such as interaural cross correlation and lateral energy fraction, and the sound strength, or gain, are of growing importance in representing the quality of experience and preference among audience members, musicians, and actors alike.

References

1. A. Cocchi et al., *Proceedings of the 6th International FASE Congress of the Federation of Acoustical Societies of Europe* (1992), p. 25; A. Farina et al., *Proceedings of the 14th International Congress on Acoustics*, L. Piezi, ed. (1992), paper F5-7.
2. B. F. G. Katz, *Acoust. Res. Lett. Online* **5**, 158 (2004).
3. B. N. J. Postma, B. F. G. Katz, *Virtual Real.* **19**, 161 (2015); *J. Acoust. Soc. Am.* **140**, 4326 (2016).
4. L. Tronchin, A. Farina, *J. Audio Eng. Soc.* **45**, 1051 (1997).
5. B. N. J. Postma, B. F. G. Katz, in *Proceedings of the 22th International Congress on Acoustics*, F. Miyara et al., eds., Argentinian Acousticians Association (2016), paper ICA2016-269.

Brian F. G. Katz
(brian.katz@sorbonne-universite.fr)
Sorbonne University
Paris, France

Damian Murphy
(damian.murphy@york.ac.uk)
University of York
York, UK

Angelo Farina
(farina@unipr.it)
University of Parma
Parma, Italy

Life and signs of the Casimir effect

We would like to offer a few comments in connection with the article "Science and technology of the Casimir effect" by Alex Stange, David Campbell, and David Bishop (Physics Today, January 2021, page 42). First, as Steve Lamoreaux mentioned in another excellent article on the Casimir effect (Physics Today, February 2007, page 40), Niels Bohr played a brief but seminal role in Hendrik Casimir's thinking. With Dirk Polder, Casimir calculated the large-distance van der Waals interaction without reference to zero-point energy.

In a 1992 letter to one of us (Milonni), Casimir recalled mentioning his results to Bohr during a walk sometime around 1947. When Casimir said that he was "puzzled by the extremely simple form of the expressions for the interaction at very large distance," Bohr mumbled something about zero-point energy. "That was all," Casimir wrote, "but it put me on a new track." That track led Casimir to use the zero-point electromagnetic energy of the modes of a resonant cavity to calculate the force between conducting plates. In his letter, Casimir said that he was "somewhat familiar with the theory of modes of resonant cavities and their perturbations" because of his position at the Philips Research Laboratories in the Netherlands.

Casimir remarked in a 1948 paper that the force between the plates "may be interpreted as a zero point pressure of electromagnetic waves,"[1] an interpretation fully supported by a calculation of the vacuum stress tensor.[2] That perspective might suggest, as do Stange and his coauthors, that the net inward pressure results from a "higher density of modes outside the plates" than inside. But such an argument is superficial in that the calculated inward and outward forces on the plates both diverge. In fact, the spectral mode density of the field between the plates can be greater at some frequencies than it is outside the plates. And it depends, of course, on the boundary conditions for the electric and magnetic fields.[3]

Stange and his coauthors highlight the major role Casimir forces play in microelectromechanical systems (MEMS) today. Interestingly, when one of us (Maclay) and two coauthors tried in 1994 to publish the first paper on the potential role of quantum forces in MEMS,[4] the reviewers initially rejected it on the grounds that the dimensions of MEMS, typically in the tens or hundreds of microns, made the discussion irrelevant.

Stange and his coauthors describe how repulsive Casimir forces can result from different dielectric properties of the interacting objects. Repulsive Casimir forces can also arise from combinations of dielectric and permeable materials, as shown in 1974 by Timothy Boyer. When one of two parallel plates is a perfect conductor and the other is infinitely permeable, for example, the force between them is repulsive. And whether the Casimir force is attractive or repulsive generally depends on the geometrical configuration of the interacting bodies. The Casimir force on a perfectly conducting sphere, for example, is repulsive, in contrast to Casimir's assumption that it should be attractive. More recently, researchers have focused on the possibility of realizing repulsive Casimir forces with metamaterials and chiral media. Qing-Dong Jiang and Frank Wilczek, for instance, have shown that chirality can be employed to obtain Casimir forces that are "repulsive," "enhanced," and "tunable."[5]

References

1. H. B. G. Casimir, *Proc. K. Ned. Akad. Wet.* **51**, 793 (1948), p. 795.
2. L. S. Brown, G. J. Maclay, *Phys. Rev.* **184**, 1272 (1969).
3. See, for instance, P. W. Milonni, R. J. Cook, M. E. Goggin, *Phys. Rev. A* **38**, 1621 (1988); V. Hushwater, *Am. J. Phys.* **65**, 381 (1997); and references therein.
4. F. M. Serry, D. Walliser, G. J. Maclay, *J. Microelectromech. Syst.* **4**, 193 (1995).
5. Q.-D. Jiang, F. Wilczek, *Phys. Rev. B* **99**, 125403 (2019).

Jordan Maclay
(jordanmaclay@quantumfields.com)
Quantum Fields LLC
St Charles, Illinois

Peter Milonni
(peter_milonni@comcast.net)
Los Alamos National Laboratory
Los Alamos, New Mexico

Algebra-based high school physics

About 60% of all physics majors chose the field because of their positive experience in high school physics courses, according to a study by the American Institute of Physics presented at the 2018 summer meeting of the American Association of Physics Teachers. In most US public high schools, only the upper 35% of the student body is allowed or encouraged to enroll in algebra-based physics classes. Such restrictions often exclude many otherwise bright and capable youngsters—of all races and ethnicities—from that formative first step toward a future in physics.

At one time, weakness in elementary mathematics, including fractions, decimals, and long division, and in basic algebra prevented even higher-level students from being successful in algebra- and trigonometry-based high school physics courses. Those mathematics deficiencies also have prompted a number of public high schools to discontinue entire physics programs. As a certified high school physics and chemistry teacher who taught mostly in Chicago inner-city high schools, where a majority of students were Black or Hispanic, from 1975 to 2001, I initially found a severe deficiency in elementary arithmetic skills. However, inexpensive electronic hand calculators have made it possible for many students to properly solve basic algebra problems that are common in physics and chemistry.

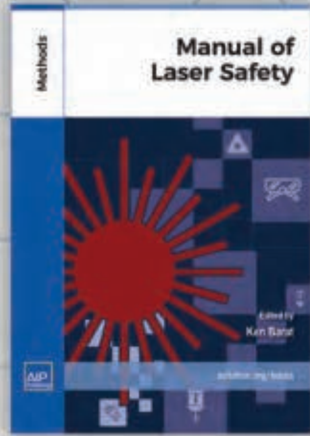
Drills and practice with standard physics formulas helped my physics students remember their first-year algebra, if not learn those algebra skills in the first place. Many high school algebra-based physics texts have only a few problems of the same type, which often does not provide good learning experiences for students of any level. To enhance their learning, I would have students do about 10 physics problems per formula; I would go around the room and show the students how to do the first problem and then allow them to solve three more of the same type using the same formula. By solving for all the formula's variables in the same manner, students learn or reinforce their algebra skills as they master their scientific problem-solving. After I retired I learned that Black students I had taught at Carver Area High School and Paul Robeson High School who entered nearby Chicago State University passed their university chemistry courses, and some even graduated with degrees in chemistry.

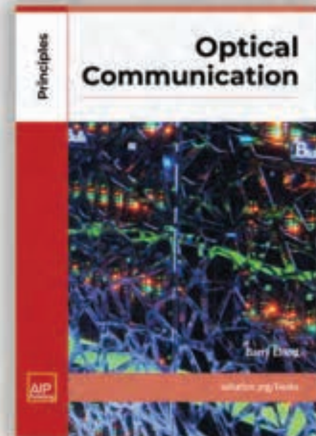
I believe that inner-city Black and Hispanic students have great potential for majoring in physics and chemistry in higher education. About two-thirds of US high schools do not have a teacher who majored in physics or physics education.[1] Also, collegiate physics programs are being eliminated because of low enrollments. If high schools open algebra-based physics to all students and provide extra physics help—including mathematics remediation, which is done everywhere in the world except in the US—not only will many student lives be enhanced, but more students will major in physics, thereby helping rescue higher-education programs at risk of discontinuation, and there will be many more competent, certified high school physics and chemistry teachers.

Reference

1. S. White, J. Tyler, *Focus On: Who Teaches High School Physics?*, American Institute of Physics (2014).

Stewart Brekke
(stewabruk@aol.com)
Downers Grove, Illinois


Methods
Manual of Laser Safety
Ken Barat


Principles
Optical Communication


Perspectives
Photovoltaic Sustainability and Management


Principles
Light Sheet Microscopy and Imaging
COMING SOON
Partha Pratim Mondal

When does a molecule make up its mind?

Speedier laser measurements are unlocking long-sought secrets of chemical dynamics.

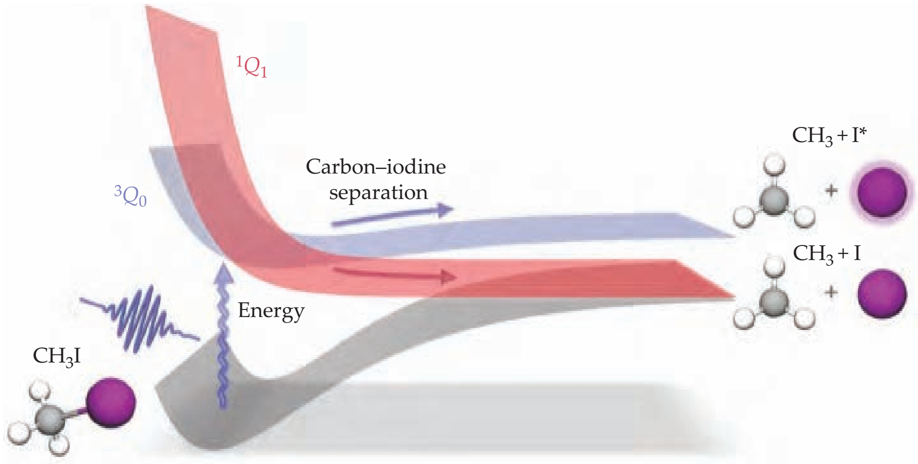


FIGURE 1. METHYL IODIDE (CH_3I), when excited from its ground state (gray) into the $^3Q_0$ state (blue), quickly encounters a choice. One-quarter of the molecules cross into the $^1Q_1$ state (pink), which dissociates into a methyl radical (CH_3) plus a ground-state iodine atom; the rest stay in the $^3Q_0$ state, which yields CH_3 plus excited-state I*. Although the outcome is well known, experiments on the ultrafast time scale of the molecular decision are needed to uncover the details of the dynamics. (Courtesy of Kristina Chang.)

When a molecule reaches a point in its configuration space where the energies of two quantum states cross, it has a choice to make: It can continue along in the state it was in, or it can cross over to the other one. Understanding what happens at those crossroads—known as conical intersections—has been a longstanding challenge in chemical physics. In some molecular systems, most molecules make the switch; in others, most don't. And there's no reliable means of predicting which systems are which.

The problem is that everything happens so fast. Electrons and nuclei alike rearrange rapidly at a conical intersection. Their motions become strongly coupled, and the Born–Oppenheimer approximation—the principle that nuclear and electronic degrees of freedom can be treated separately—breaks down. Because the Born–Oppenheimer approximation is the foundation of nearly all conventional quantum chemical calculations, theorists are left with few tools for describing dynamics at conical intersections. And experimental studies on such rapid time scales aren't much easier.

Now, Kristina Chang of the University of California, Berkeley, her advisers Daniel Neumark and Stephen Leone, and their colleagues have developed a system of ultrafast lasers that can directly observe molecules passing through a conical intersection.[1] In an experiment on the widely studied prototype molecule methyl iodide (CH_3I), they measured the time scale of the molecule's choice. Their results agree well with what theorists predicted for the system two years ago[2]—a promising sign that theorists and experimenters alike are well on their way to understanding molecular decisions.

Life choices

Conical intersections are thought to be critical to many of the ways in which molecules convert or make use of energy, especially in living things. Examples include photosynthesis, human vision, and DNA's mechanism for resisting UV damage.

Nature has evolved a genetic code consisting of just four DNA bases: adenine, cytosine, guanine, and thymine. Plenty of other molecules exist that could fit just as easily inside DNA's double helix. For example, 2-aminopurine is structurally identical to adenine except for the placement of an NH_2 group, and it pairs with thymine just as well as adenine does. Nature's choice of adenine over 2-aminopurine seems to have come down to how the molecules respond to UV light.

When adenine absorbs a UV photon—a frequent occurrence in our sunlit world—it returns to its ground state within a picosecond; 2-aminopurine, on the other hand, takes tens of thousands of times as long.[3] The longer a molecule remains in an excited state, the more opportunity the excitation energy has to initiate a reaction that can lead to a genetic mutation. In fact, biologists sometimes use 2-aminopurine as a substitute for adenine in experiments when they want to induce mutations deliberately.

Adenine's short excited-state lifetime is attributable to a conical intersection that efficiently funnels molecules from the excited state to the ground state. It's not much of an exaggeration, therefore, to say that life on Earth as we know it owes its existence to conical intersections.

But just because a conical intersection can funnel molecules between states doesn't mean it always does. After all, a molecule can pass through a conical intersection and remain in the same state where it started. A more complete understanding of conical intersections is needed to explain why biomolecules behave the way they do. But because biomolecules themselves have many cumbersome degrees of freedom, researchers studying conical intersections often focus on CH_3I, one of the simplest organic molecules with a mere four chemical bonds, one of them different from the rest.

Speedy spectroscopy

The basic dynamics of CH_3I are well known. When the molecule is excited with UV light at around 277 nm, the carbon–iodine bond invariably breaks. Of the newly freed I atoms, about one-fourth are formed in their ground state, and the remaining three-fourths are in an excited state I*. (Any excess energy goes into either CH_3 vibrations and rotations or the kinetic energy of the two fragments as they fly apart.)

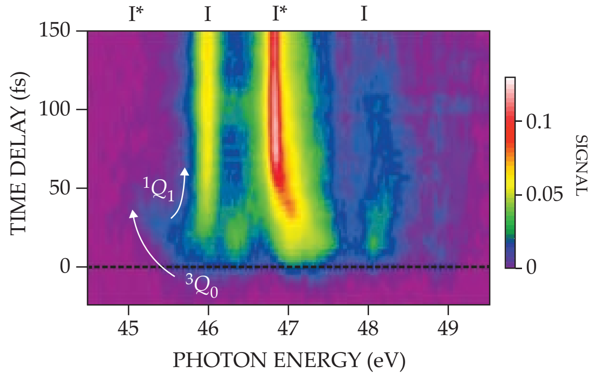


FIGURE 2. ATTOSECOND TRANSIENT absorption spectroscopy shows two distinct excitations of dissociating methyl iodide molecules. Although the 45–46 eV excitation quickly becomes quantum mechanically forbidden for molecules remaining in the $^3Q_0$ state, the split with the $^1Q_1$ state is still visible. (Adapted from ref. 1.)

Figure 1 depicts what's going on. The excitation takes the molecules into the $^3Q_0$ state, right on the brink of a conical intersection with the $^1Q_1$ state. Molecules that remain in $^3Q_0$ yield I* atoms; those that cross into $^1Q_1$ yield ground-state I.

Neumark's and Leone's labs specialize in ultrafast chemistry, the study of the rearrangements of atomic nuclei on time scales of femtoseconds (see PHYSICS TODAY, December 1999, page 19) and motions of electrons on time scales of attoseconds (see PHYSICS TODAY, April 2003, page 27). In 2014 they and then-student Annelise Beck developed the technique of attosecond transient absorption spectroscopy for probing rapid molecular movements.[4]

Many other attosecond techniques can be complicated to interpret (see, for example, PHYSICS TODAY, January 2018, page 18). In contrast, transient absorption spectroscopy is much like any other form of absorption spectroscopy: By detecting which energies of incident light a sample absorbs, researchers infer what molecular states it contains. Because the incident light is a pulse that lasts just attoseconds, the measurement can be made with exquisite temporal precision.

A study of the CH_3I conical intersection proceeds as follows. An initial UV pulse excites a population of molecules into the $^3Q_0$ state to start them on their journey to carbon–iodine dissociation. After a tunable time delay of tens to hundreds of femtoseconds, the dissociating molecules are blasted with a broadband attosecond pulse that probes their states. If all the molecules are still in the same state, they haven't yet passed through the conical intersection. But if they're split into two subpopulations of different energies, then they have.

That experiment has been performed before,[5] but it didn't observe the molecules passing through the conical intersection. The sticking point was the initial UV pump pulse. Because of the combination of power and wavelength needed to excite the CH_3I molecules, it wasn't possible to make the pulses any shorter than 100 fs, far longer than the time scale the researchers were hoping to measure. An experiment's time resolution isn't strictly limited by its pulse duration—it's possible to use 100 fs pulses to measure times somewhat shorter than 100 fs—but the passage through the conical intersection was still too fast to be meaningfully detected.

Time to split

"Short, intense UV pulses are difficult to generate," explains Chang, "because we have to make them via the up-conversion of longer wavelengths. The conversion process is often inefficient, so we get UV pulses that are too low in energy to do any kind of useful measurement."

Chang and her colleagues addressed the challenge head-on by developing a new optical setup that made the frequency conversion more efficient. With a free-standing crystal of the nonlinear optical material beta barium borate, they frequency-tripled an ultrashort near-IR pulse. The result: a UV pulse with sufficient power to excite a CH_3I population, but with a duration of just 20 fs.

The resulting attosecond transient absorption spectroscopy data are shown in figure 2. The attosecond pulse excites two distinct transitions in the dissociating molecules, one in the range of 45–46 eV, the other around 47–48 eV. By the time the dissociation is complete—that is, when the pulse is exciting the unbound I atoms—the lower-energy transition on the I* state is quantum mechanically forbidden.

But prior to about 50 fs, a weak blue streak veering off to the left, from molecules remaining in the $^3Q_0$ state, is visible, as is the much stronger yellow streak from those crossing into the $^1Q_1$ state. By fitting the data, the researchers pinpoint the onset of the splitting to 15 ± 4 fs, in good agreement with the theorists' prediction[2] of 13 fs.

"Fifteen femtoseconds is less than a fifth of the carbon–iodine dissociation time of 80–100 fs," says Chang. "That means the molecule makes its 'decision' early on in the reaction of whether to switch states at the conical intersection or not." That conclusion might seem obvious from the way figure 1 is drawn—if the molecules are excited so close to the conical intersection, they should have no choice but to pass through it immediately—but that detail of the system had never before been confirmed experimentally.

The next step is to repeat the experiment with other molecules to see how conical-intersection passage times vary with chemical structure. When iodine is part of a bulkier organic molecule, for example, it dissociates in a similar way and passes through a similar conical intersection, but the outcome varies significantly. Only one-fourth of CH_3I molecules switch states at the conical intersection and produce ground-state I atoms. But in isopropyl iodide (which contains three carbon atoms), two-thirds of molecules make the switch, and in tert-butyl iodide (with four carbon atoms), more than 90% do.[6]

Could that trend be connected to an observable difference in the dynamics of how the molecules pass through the conical intersection? Ultimately, after all, chemical physicists are interested not just in when molecules make up their minds but in what they decide and why. By systematically studying the relationships between structure, dynamics, and outcome, Chang and colleagues hope to work toward an understanding of how molecular choices influence both chemistry and biology.

Johanna Miller

References

1. K. F. Chang et al., *J. Chem. Phys.* **154**, 234301 (2021).
2. H. Wang, M. Odelius, D. Prendergast, *J. Chem. Phys.* **151**, 124106 (2019).
3. H. Kang et al., *J. Am. Chem. Soc.* **124**, 12958 (2002).
4. A. R. Beck, D. M. Neumark, S. R. Leone, *Chem. Phys. Lett.* **624**, 119 (2015).
5. A. R. Attar, A. Bhattacherjee, S. R. Leone, *J. Phys. Chem. Lett.* **6**, 5072 (2015); L. Drescher et al., *J. Chem. Phys.* **145**, 011101 (2016).
6. K. F. Chang et al., *Nat. Commun.* **11**, 4042 (2020).

Viscoelastic fluids with no strings attached

Dispensing a fluid is quick and clean when the nozzle is rotated. The fluid's elastic properties are the reason why.

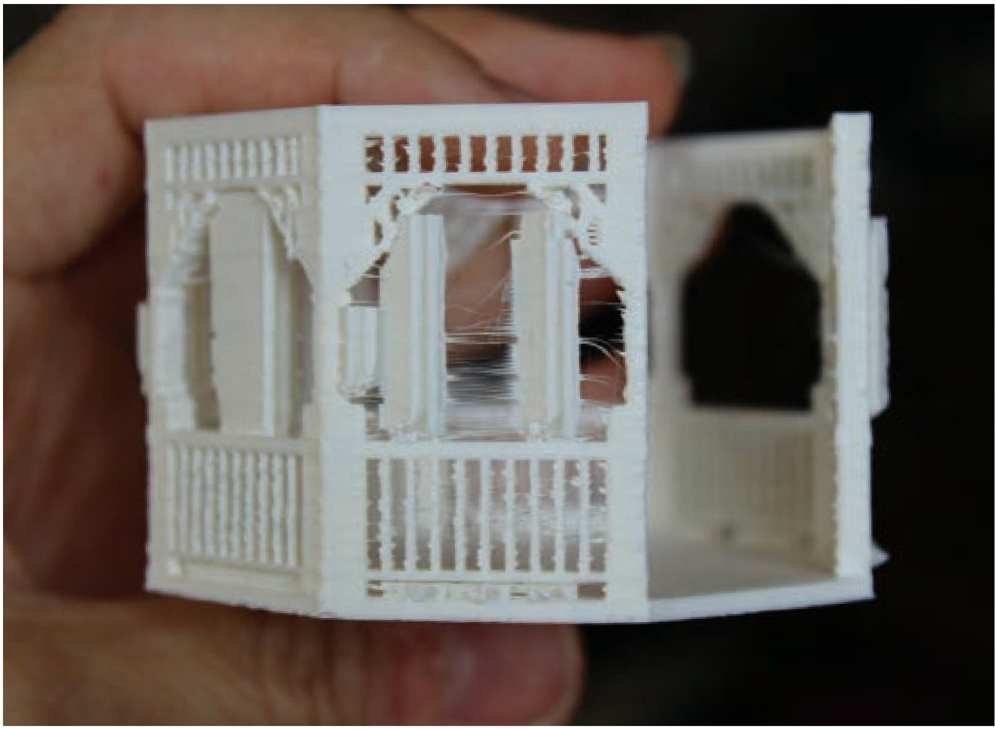


FIGURE 1. PLASTIC STRINGS mar the 3D-printed gazebo shown here. The strings appear when the printer nozzle lifts to detach from the deposited viscoelastic fluid. That retraction elongates the connecting fluid bridge, and after detachment, the strand sticks out. (Photo by Vicky Somma, CC BY-NC-SA 2.0.)

If you've ever wielded a glue gun, you likely dealt with the wisps of glue that trailed after it. The same issue plagues additive manufacturing: Instead of a tidy reproduction of the desired shape, a three-dimensional printer constructs an object marred by plastic strings, as shown in figure 1. Those strings are difficult to prevent when dispensing plastics, polymers, and other viscoelastic fluids, which behave as viscous fluids at low speeds and as elastic solids at high speeds.

When an ordinary Newtonian fluid such as water is dispensed from above, it bridges the gap between the target substrate and the nozzle. If the gap is kept below a critical value, the connection is stable. At or above that value, gravity gradually drains the liquid until the bridge breaks. To speed up the severance, one can simply lift the nozzle to thin the bridge until it splits.

Retraction also expedites the breakup of viscoelastic liquid bridges. But as the bridge is elongated, the liquid becomes more elastic and harder to break. Once the connection snaps, elongated strands and droplets, shown in figure 2a, can fall in unpredictable and potentially troublesome ways. For example, in the process of gluing electronic components on printed circuit boards, a stray string of glue can ruin the device. In short, there's a tradeoff between dispensing quickly and dispensing cleanly.

Now Amy Shen and Simon Haward, of Okinawa Institute of Science and Technology Graduate University in Japan, and their colleagues have developed a new rotation-based method to dispense viscoelastic fluids. Their high-speed video analysis and viscoelastic-flow simulations show that twisting the fluid bridge, rather than stretching it, quickly breaks the connection without producing strings.[1]

A new twist

The problem of how to efficiently dispense viscoelastic fluids first came to the attention of Hammad Faizi, a co-lead author on the study, in 2015. An undergraduate at the University of Hong Kong at the time, he learned about the issue in a summer internship with ASM Pacific Technology, a company that specializes in designing and manufacturing semiconductor technology. When Faizi returned to school, he and fellow undergraduate San To Chan delved into the problem under the supervision of Anderson Shum. They found that rotating the dispensing nozzle yielded cleaner and faster application, a discovery that led to a patent application and their undergraduate dissertations. But they couldn't explain the underlying physics.

To pursue his PhD, Chan joined Shen's group, which specializes in the rheology of viscoelastic fluids. And in 2019 Shen, Chan, and their colleagues resumed investigating the effects of rotation on fluids. Chan built a simple experimental setup composed of two plates: an upper plate that can rotate and a lower plate that can move vertically to set the distance separating them. The researchers sandwiched 40 μL of liquid between the plates, separated by 2 to 4 mm, to form the expected liquid bridge with a radius on the order of a millimeter. They tested four viscoelastic silicone oils with different viscosities and, for comparison, a Newtonian fluid called Infineum S1054.

At various plate separations and rotation speeds, Shen and her group members filmed the fluid to track the bridge's radius as a function of time and analyzed the data with a program written by Faizi. They found that as both silicone and Infineum are twisted, the top half of the fluid rotates like a solid object, and the lower half is nearly stationary. So the twisting localizes at the neck where the top and bottom halves meet, similar to that of a washcloth being wrung out, and pinches off the liquid bridge. When the plate rotates at 35.3 Hz, Infineum takes about 9 s to break, compared with about 60 s with no rotation. The silicone bridge also breaks faster under rotation—in fact, even an otherwise stable bridge snaps apart in around 1 s. But instead of the broad concave form of the Infineum bridge, the thinning silicone bridge has a peculiar narrow indentation on its sides.

The dynamics of viscoelastic bridges depends on a balance of inertial, elastic, capillary, and gravitational stresses and a host of parameters that includes density, surface tension, viscosity, elasticity, plate separation, and rotation rate. But surprisingly, the results are straightforward: The silicone bridge radius decreases with time t according to the power law $R \propto t^{-\beta}$, where β depends on a single dimensionless parameter, the Tanner number, which characterizes the relative importance of the torsion-induced elastic stresses compared with capillary stresses. The Tanner number is similar to the Reynolds number, which characterizes the relative importance of viscous and inertial forces.

As the viscoelastic bridge is twisted, the shear stresses at the neck generate stresses perpendicular to the fluid's surface, which doesn't happen in ordinary Newtonian fluids. Those shear-induced normal stresses overcome surface tension to form an indentation in the side of the bridge that's tenths of a millimeter wide. The indentation further localizes the shear stress and thus further increases the normal forces, which draw the indentation toward the middle of the bridge. That feedback loop results in the indentation propagating inward until the upper and lower halves of the liquid are separated, without any stray strands, as shown in figure 2b.

Because the indentation splits an otherwise stable liquid bridge, gravity can't be the source of the behavior. And given that the Newtonian fluid doesn't form an indentation, the behavior must be related to an elastic effect. So Shen and her colleagues hunted for an elastic instability that could explain their observations.

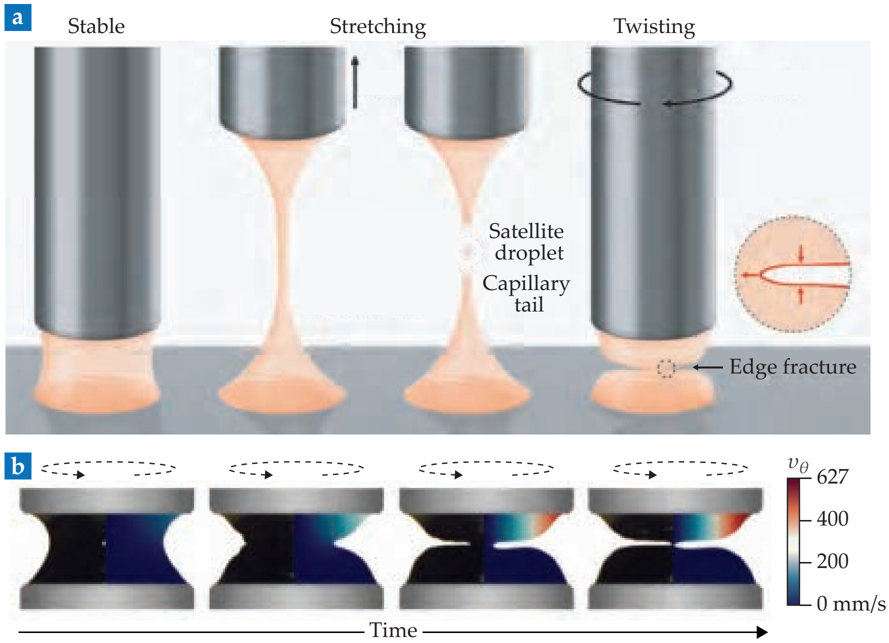


FIGURE 2. VISCOELASTIC FLUIDS between two surfaces. **(a)** Below a critical nozzle height (left), a fluid (pink) forms a stable connection between the target substrate and the nozzle. When the nozzle is above a critical height (center), the fluid breaks and forms undesirable satellite droplets and strands known as capillary tails. If, instead, the nozzle rotates (right), an indentation formed by the normal stresses (red arrows), shown in the inset, cleanly severs the fluid bridge. **(b)** Experimental photos of silicone oil rotated at 35.3 Hz (left half of each image) and simulated azimuthal velocity v_θ (right halves) show how the indentation forms in the side of the bridge and propagates to the center. (Adapted from ref. 1.)

On the edge

Chan noticed that the indentation behavior resembles a flow instability first reported in 1963: edge fracture.[2] In that phenomenon—which emerges in various viscoelastic fluids, including toothpaste—shearing above a critical rate induces a sudden indentation, which localizes the shear stress and invades the fluid.

Edge fracture looks similar to what Shen and her group observed, but the phenomenon was previously investigated in a regime where the radius of the fluid was an order of magnitude larger than the height. Could the same behavior occur in their liquid bridge, with its roughly equal radius and height? To verify the connection, the group teamed up with Patrick Anderson and his graduate student Frank van Berlo of Eindhoven University of Technology in the Netherlands to perform calculations complementing the experiment.

In his simulations, van Berlo had to verify that the liquid bridge system possessed properties characteristic of edge fracture. For example, the indentation should have a normal stress tugging its tip toward the center and normal stresses squeezing its upper and lower surfaces, as shown in the inset of figure 2a. Tweaking the right set of inputs to test the relationship between the normal and shear forces and the formation of indentations, van Berlo confirmed edge fracture as the origin of his colleagues' observations. Their study thus expanded the parameter space of such research.

Edge fracture is typically viewed as a problem in rheological measurements because it makes results harder to interpret. But for clean dispensing, edge fracture is advantageous. "Instead of avoiding the elastic property of the viscoelastic liquid bridge, we amplify it to the extreme where fracture occurs," says Shen.

The simplicity of the researchers' twisting technique means it's feasible in industrial settings. One potential application is food engineering: The rotating nozzle could neatly apply jam to sugar cookies on a conveyer belt, for example.

But some fundamental questions remain. Although researchers now know the mechanism behind the breakup, it isn't clear why the radius decays according to a power law when rotated. Shen and her collaborators plan to partner with a theorist to further explore the issue and to derive an analytical equation for the decay. Such an equation may provide a route to novel methods for measuring rheological properties of viscoelastic fluids, such as shear-induced normal stresses, that have been difficult to probe by traditional methods.

Heather M. Hill

References

1. S. T. Chan et al., *Proc. Natl. Acad. Sci. USA* **118**, e2104790118 (2021).
2. J. F. Hutton, *Nature* **200**, 646 (1963).

Squeezed hydrogen and helium don't mix

The results of a high-pressure experiment that re-creates the temperatures and pressures inside Jupiter and Saturn reconcile disparate space-based observations.

By mass, the giant gaseous planets Jupiter and Saturn consist of about 75–85% hydrogen and helium. (Uranus and Neptune, by contrast, are ice giants and contain far greater proportions of ammonia, methane, and water.) Researchers have hypothesized that the remaining composition of Jupiter and Saturn could be a few other gases, ice, and rock. Determining what those exact fractions are and how the planetary structures evolved over time depends critically on the H–He equation of state.

On the surface of gas giants, hydrogen and helium form a homogeneously mixed layer. In the 1970s, physicists predicted that at the high temperatures and pressures inside gas giants, the two lightest elements may separate and form a region of demixing, or immiscibility.[1] Today scientists know that at some depth in the planet's interior, molecular hydrogen dissociates and ionizes to form metallic fluid hydrogen. The helium is expelled from the mixture and is predicted to rain down from the immiscibility region to another layer, in which the conditions allow hydrogen and helium to mix again. Jupiter's central core is suspected to be solid.

Since that first prediction, many scientists have sought to better describe where in the planet's interior the H–He phase separation occurs and how gas giants are structured. Condensed-matter models based on density functional theory, for example, begin from the basic principles of quantum mechanics to approximate the system's electron density (see the article by Andrew Zangwill, PHYSICS TODAY, July 2015, page 34).

Despite such efforts, uncertainties remain. Models of H–He mixing in planetary interiors are highly dependent on the particular functional form a researcher chooses, and no model reliably describes the range of experimental results. Another challenge is accounting for the entropy within a mixed system of hydrogen and helium, which makes the calculations difficult.

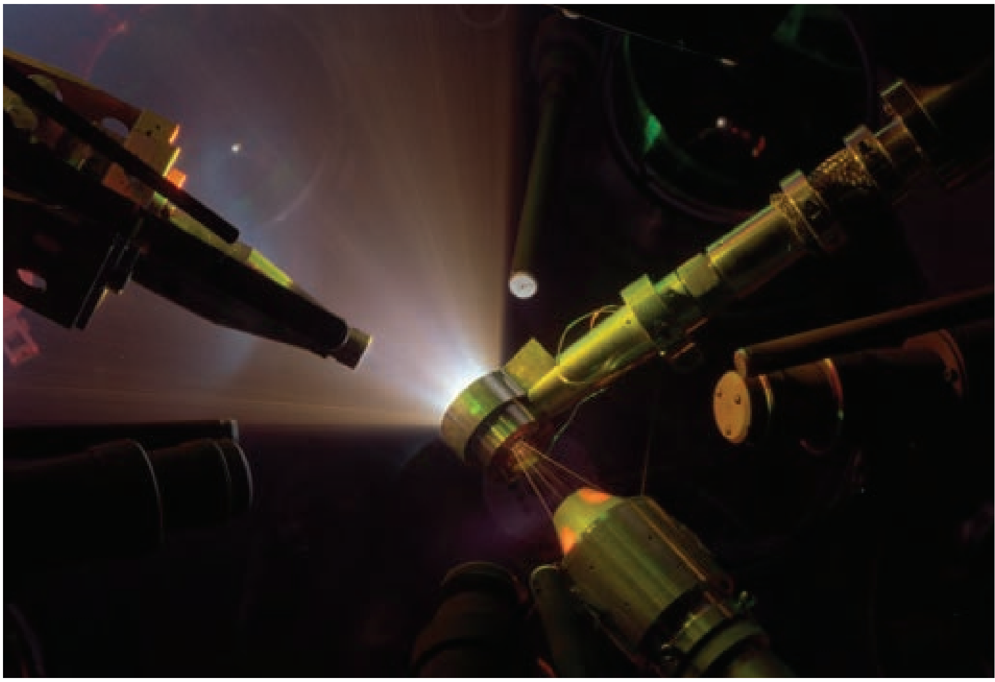


FIGURE 1. COMPRESSION EXPERIMENTS. At the University of Rochester's Laboratory for Laser Energetics, researchers uncovered new phase behavior in a mixture of hydrogen and helium gas. The cylindrical diamond anvil cell in the center of the photo compressed the gas sample statically before a laser-induced shock wave propagated through the sample and squeezed it dynamically to the pressure and temperature range of Jupiter's interior. (Photo courtesy of OMEGA Laser Facility.)

Laboratory measurements are also challenging. Researchers studying H–He mixing with standard static compression methods using a diamond anvil cell must contend with the hydrogen's reactivity with the diamonds. On the other hand, dynamic compression with a laser-induced shock wave demands an initially homogeneous H–He mixture with a density three times that of cryo-liquid hydrogen to span the pressure conditions of a planetary interior.

Fifteen years ago, Stephanie Brygoo and Paul Loubeyre of the French Alternative Energies and Atomic Energy Commission; Raymond Jeanloz of the University of California, Berkeley; and their colleagues began blending the two types of high-pressure experiments. Their recent results not only support the existence of an immiscibility region in Jupiter but also provide experimental evidence for a four-layered planetary structure that's consistent with indirect spacecraft observations.[2]

Shining hydrogen

Jupiter's interior temperatures span thousands of kelvin, and its pressures span hundreds of gigapascals, about 1 million to 40 million times that of Earth's atmosphere. To reach such high pressures and temperatures in the lab, Brygoo and colleagues subjected samples of homogeneously mixed hydrogen and helium to static compression followed by dynamic compression. The researchers modified the apparatus so that one diamond window was only 300 μm thick, far thinner than the typically millimeters-thick walls of the anvil.

In the experiment, a diamond anvil cell first squeezed the samples to 4 GPa. The modifications allowed a laser-induced shock wave to propagate through the diamond window without deteriorating. Using the OMEGA laser at the University of Rochester's Laboratory for Laser Energetics, the researchers compressed the samples further, to pressures up to about 200 GPa.

Figure 1 shows the experimental apparatus at the moment that the laser shock wave propagated through the hydrogen and helium sample, which was squeezed first by the diamond anvil cell. Sébastien Hamel of Lawrence Livermore National Laboratory says that "compressing hydrogen to such a degree without cracking the diamonds used to apply pressure, combining the static compression with laser-based shock waves to reach the required conditions, and building the diagnostic tools necessary to make the observations were incredible challenges that this team overcame."

During the experiment, Brygoo and her colleagues observed a sudden change in reflectivity of the laser light bouncing off the H–He mixture as the shock front passed through the sample. *Ab initio* calculations and theory predict that as the H–He mixture becomes conductive, insulating helium separates from metallic hydrogen. The effect increases reflectivity, a measurable signature of the phase separation. Loubeyre says that he and his colleagues "were not expecting such a clear discontinuity. It was expected to be smaller than what we observed."

Figure 2a shows the H–He mixture's phase diagram with the predicted demixed region colored in pink. The data, which are similar to previously published theoretical predictions,[3] indicate that the samples reached the temperature–pressure conditions of Jupiter's interior as the shock wave propagated through the system. "Phase separation has started rather recently in Jupiter, so the distribution of helium is not completely perturbed," says Loubeyre. Therefore, if the helium distribution is constant, the researchers estimate that about 15% of Jupiter's radius—the pink shaded region in figure 2b—forms an immiscibility region.

Reconciling observations and theory

The new laboratory results help tie together different spacecraft observations of Jupiter. In 1995, NASA's *Galileo* probe collected measurements of the helium abundance in Jupiter's atmosphere and found it smaller than the presolar value—the abundance thought to be present at the formation of the solar system.[4]

NASA's *Juno* probe, however, recorded measurements of a high gravitational moment in 2016 that indicated a model discrepancy.[5] The observations suggest that the planet's interior density should correspond to a high helium abundance in the outer layer. To reconcile the *Galileo* observations with the *Juno* measurements, modelers had to make modifications to the H–He equation of state.

The results reported by Brygoo and her colleagues appear to resolve the issue by providing experimental support for a Jupiter with four distinct layers. The immiscibility region lies between the homogeneously mixed H–He outer layer observed by *Galileo* and the deeper, well-mixed metallic H–He layer compatible with the *Juno* measurements. A fourth layer—a dense core of mostly heavy elements—lies at the center of the planet.

According to planetary scientist Christopher Mankovich of Caltech, observing the separation of hydrogen and helium "is a pretty big deal. The existence of this immiscibility region is an approximately 40-year-old idea that has shaped much of our thinking about the structures of Jupiter and Saturn. Finally there is a modicum of compelling experimental evidence."

Alex Lopatka

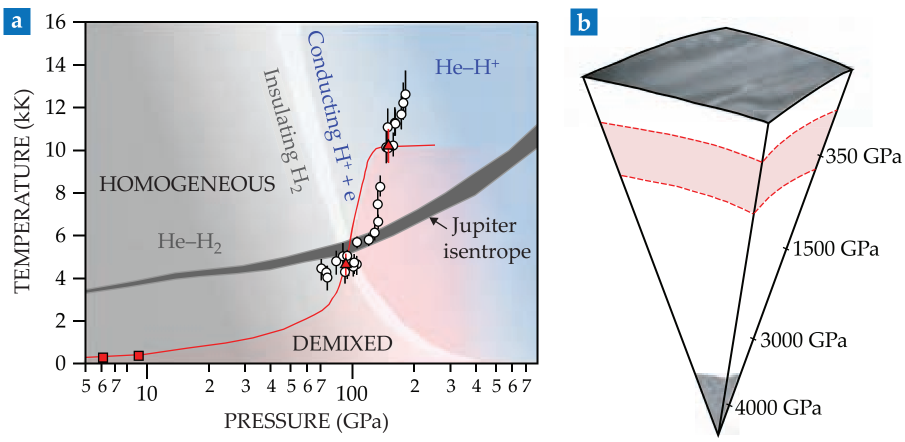


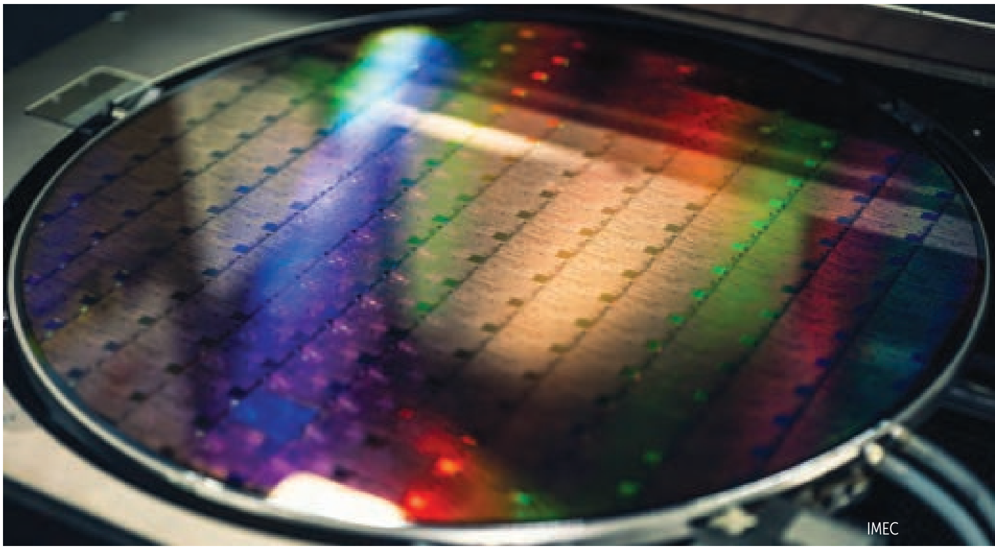
FIGURE 2. JUPITER'S IMMISCIBILITY REGION. (a) Previously collected reflectivity data from static compression experiments (red squares) and new dynamic compression measurements (red triangles) suggest a boundary (red line) between the predicted homogeneous and demixed regions of the H–He phase diagram. The white circles indicate that the samples span a region of phase space where hydrogen and helium separate. That Jupiter's temperature–pressure conditions cross the demixed region means that an immiscibility layer likely exists in the planet's interior. **(b)** With a constant helium distribution, the immiscibility layer (pink shading) spans about 15% of the planet's total radius. (Adapted from ref. 2.)

References

1. D. J. Stevenson, E. E. Salpeter, *Astrophys. J., Suppl. Ser.* **35**, 221 (1977).
2. S. Brygoo et al., *Nature* **593**, 517 (2021).
3. S. Hamel, M. A. Morales, E. Schwegler, *Phys. Rev. B* **84**, 165110 (2011); M. A. Morales et al., *Phys. Rev. B* **87**, 174105 (2013).
4. U. von Zahn, D. M. Hunten, *Science* **272**, 849 (1996).
5. S. J. Bolton et al., *Science* **356**, 821 (2017). PT

Europe's experiment in funding graphene research is paying off

The outputs of the European Union's eight-year-old Graphene Flagship include products, spinoffs, and "Spearheads."



IMEC

A GRAPHENE LAYER is deposited onto a silicon wafer for microelectronics and computing applications developed at Imec, the nonprofit Belgian nanoelectronics and digital technologies research institute.

In 2013 the European Union (EU) embarked on one of its three largest-ever targeted technology R&D programs, a 10-year, €1 billion ($1.18 billion) program called the Graphene Flagship, with the aim of creating applications and markets for the two-dimensional material and establishing Europe at the forefront of the technology. Ultimately, its impact and whether it has been worth the public spending and effort won't be known for years to come.

"Over the past seven years, the Graphene Flagship has successfully brought graphene out of the lab, creating a fruitful European industrial ecosystem that develops applications of graphene and layered materials," says the flagship's 2020 annual report. "Today, our industrial family includes over 100 companies working together with the Graphene Flagship's academic partners in fields ranging from the automotive and aviation industries to electronics, energy, composites and biomedicine."

The EU launched another flagship, focused on the human brain, in 2013 (see PHYSICS TODAY, December 2013, page 20). A third flagship, covering quantum technologies, got underway in 2018. Both are also 10-year, €1 billion efforts.

With two and a half years remaining, the Graphene Flagship claims to have spawned roughly 90 products that incorporate graphene in some way, with applications as disparate as Hall effect sensors that have 10 times as much sensitivity as their silicon-based counterparts and earphones with enhanced treble and bass. Other niche products include a bicycle tire with improved traction, a tennis racket with superior flexibility and durability, a motorcycle helmet with high-impact dispersion, and a high-performance air conditioner. Graphene chemical vapor deposition systems that are used in semiconductor applications originated from flagship R&D.

Measured by the number of patents and papers published per euro spent, "we do extremely well," relative to the outputs of academic research, says Jari Kinaret of Chalmers University of Technology in Gothenburg, Sweden. The condensed-matter physicist has directed the Graphene Flagship from its outset. But few, if any, technology-specific collaborations on similar scales exist, so evaluations of the flagship's performance are largely subjective. "I have had a standing offer for many years that if someone could come up with a measurable, meaningful key performance indicator, I will buy them a nice dinner," Kinaret says.

Kari Hjelt, the flagship's head of innovation, says a peer-reviewed assessment of its economic impacts will be performed once the program is finished. But he cautions that the scorecard will be incomplete because the time from research to commercialization typically is 15 years or longer. What is clear so far, he notes, is that the flagship has encouraged greater risk-taking on the part of European industry to venture into new product commercialization.

Unlike in the US, where industry is typically expected to put up some of its own funds for cooperative research involving government agencies, the Graphene Flagship reimburses its commercial partners for their direct R&D expenses. But the European Commission, which oversees the flagship, covers only a portion of the indirect or overhead costs that the industry partners incur, Kinaret notes.

The approximately 170 current flagship partners are about evenly divided between universities and industry. In the beginning, membership was roughly 75% academia. "But in all honesty, some of the industry partners were perhaps more spectators than players," says Kinaret. The evolution to greater industry participation was anticipated in the road map that the flagship had drawn for itself at its outset, he says.

The flagship has €150 million left to spend through 2023. About 45% of that will fund applied research, and 15% will be devoted to basic research. Research is performed in four broad topical areas: enabling science and materials; health, medicine, and sensors; electronics and photonics integration; and energy, composites, and production.

Industry-led, multiple-institution programs designed to commercialize specific graphene products—Spearhead Projects—get another 30% of the flagship budget. The remainder goes to administrative support services, which include standards development, product performance testing, and governance.

Technology push

The 11 current Spearhead Projects were initiated last year and are due to wrap up by September 2023. By design, the projects can progress only to the prototype stage, Kinaret says. A special dispensation from the European Commission is required to proceed further. Full commercialization of products occurs outside the flagship program.

Letizia Bocchi, laboratory manager of filters and medical applications at Medica SpA in Italy, directs the Graphil Spearhead Project, which is developing water filters for drinking-water taps. The filters are made of graphene oxide embedded in hollow fibers that are spun from polymers. Graphene absorbs organic contaminants, such as antibiotics and perfluorinated compounds, that would otherwise escape the mechanical filtration provided by the fibers. Among other advantages, Graphil's process doesn't produce the wastewater that reverse osmosis does.

Two other companies, France's Polymem and the UK's Icon LifeSaver, are also Graphil partners, as are Italy's national research institute, Chalmers, and the University of Manchester. A single company couldn't have afforded the required R&D, Bocchi says, and government support was essential. "We are building up a plant dedicated to production of the fiber, which is a big investment." The flagship's high visibility throughout the EU has led to many contacts and expressions of interest from other companies, she adds.

The Circuitbreakers Spearhead Project is exploiting graphene's self-lubricating properties to develop improved circuit breakers for protecting suppliers and consumers of electricity, such as wind farms, hospitals, and data centers. Grease, the existing century-old standard for ensuring the mechanical functioning of breakers, decomposes over time and must be replaced periodically. Anna Andersson, a principal scientist at Sweden's ABB, leads the project. She says the researchers are evaluating the application of graphene coatings to metal substrates by using electroplating or sintering. The other industry members are materials producers Graphmatech of Sweden and Nanesa of Italy. Chalmers, Manchester,

BEDIMENSIONAL

GRAPHENE POWDER is produced for the manufacturing and energy industries by Graphene Flagship spinoff BeDimensional, an Italian company.

AIXTRON

AIXTRON MACHINES perform chemical vapor deposition of graphene for compound semiconductor applications. The technology was developed by the German company through the Graphene Flagship.

Greece's Foundation for Research and Technology–Hellas, and the University of Rome Tor Vergata are the academic partners.

Another Spearhead, Autovision, aims to produce sensors for self-driving cars to detect objects and road curvature in dark or foggy conditions. Qurv Technologies, which leads the project, was spun out from the Spanish photonics research institution ICFO. Qurv's graphene–quantum dot image-sensor technology features wide-spectrum computer vision technology that is compatible with the CMOS manufacturing process. Veoneer, a spinoff of Autoliv, the Swedish–US company that is the world's largest supplier of automotive safety equipment, is an Autovision partner, as are the Belgian nonprofit Imec and German technology company Aixtron. If successful, the new sensors would provide an alternative to gallium arsenide and indium arsenide technologies, which require scarce metals and toxic arsenic.

A Spearhead Project, led by Airbus, seeks to develop a graphene-based thermoelectric ice-protection system for aircraft. Fiat Chrysler Automobiles (now known as Stellantis) is leading a Spearhead to produce a metal-free graphene automobile dashboard that could reduce manufacturing costs and lower fuel consumption. Another Spearhead, led by Varta Microinnovation, plans to advance a graphene–silicon anode for lithium-ion automotive batteries. Its goal is improving the lifetime of a composite that was developed in an earlier flagship phase to 1000 charge–discharge cycles.

The Graphene Flagship has already spun off a dozen companies. InBrain Neuroelectronics, formed by two flagship-partner research institutes funded by the Catalan government, is developing graphene-based implants for treatment of Parkinson's, epilepsy, and other brain disorders. In March the company announced it had raised a total of €15 million from venture capital firms and the Spanish government. BeDimensional, an Italian spinoff, develops and produces graphene for the manufacturing and energy industries.

Shoes and concrete

Many low-tech applications for graphene, such as composite materials, have moved forward without flagship support. When mixed with rubber, graphene improves the grip and durability of tires. Shoes manufactured by the UK company Inov-8 improve trail runners' purchase on slick rocks while also providing good cushioning. Researchers have found that concrete mixed with small amounts—typically 0.1%—of graphene flakes can improve the material's structural strength by 30% and thereby reduce the need for structural steel reinforcement. Adding graphene also lowers the amount of concrete required for a job and reduces the substantial carbon footprint of cement and concrete.

THE GRAPHENE FLAGSHIP'S anticipated timeline for the commercialization of various graphene applications and products. (Courtesy of Graphene Flagship.)

James Baker, who leads the University of Manchester's Graphene Engineering Innovation Centre, says the university teamed with a construction company in May to pour a local gym floor made with so-called Concretene. "Builders who have been in the business for 40 years say it's the best concrete floor they've ever seen," he says.

Despite Brexit, UK members will remain in the flagship through its conclusion. Manchester, where Andre Geim and Kostantin Novoselov discovered how to make graphene in 2004 (see PHYSICS TODAY, December 2010, page 14), participates in several of the Spearhead Projects. The university also hosts the UK's National Graphene Institute. Similar national-level graphene research organizations are located in various EU countries.

Baker says the Graphene Flagship has brought together a much larger range of universities and industry participants than Manchester could have done on its own. "With €1 billion, the flagship has got critical mass, and other activities, like toxicology, health and safety, and standards, things that are best to do collectively. It's been a key part in accelerating the whole language and supply chain around graphene, and that's critical if you're commercializing."

Terrance Barkan, executive director of the Graphene Council, a trade association based in North Carolina but with an international membership, credits the flagship for elevating awareness of the material in Europe and providing sustained support for academic R&D. "The strength of the Graphene Flagship has been providing the EU and Europe with a unified strategy in developing an industry around these advanced materials. Whereas in the US, everybody is left to their own devices," he says. But while the flagship has performed well on supporting innovation, "the commercial conversion hasn't been as big a success as one would expect." And dependence on the subsidies the flagship has provided to industry could result in products that don't sell, he says.

A tipping point

Regardless of how its performance is judged, the Graphene Flagship isn't expected to be extended beyond its 10-year term. "What happens after the flagship ends is that many of our activities will continue with the support of Horizon Europe," says Chalmers's Kinaret, referring to the EU's €95.5 billion wide-ranging R&D program that runs through 2027. "And some will have reached the level of maturity that they no longer require support of the European taxpayer."

Rather than duplicating the flagship, Barkan says, the US should drive demand by easing regulatory requirements that currently need approval from the Environmental Protection Agency for each new application of graphene. The US could also boost the growth of such products as graphene-embedded concrete by establishing low-carbon-emission requirements for road-building materials.

Forecasts for graphene demand vary widely, but there is little doubt that the market will rapidly expand. Still, graphene is being held back by a lack of application-oriented standards and uniform quality metrics. Part of the flagship program has been devoted to developing such standards, but adoption requires approval from multiple national and international standards bodies, a process that can take years. "The problem is, industry doesn't have years to wait," says Barkan, whose organization also works to accelerate the adoption of standards.

"Graphene is approaching a tipping point," says Baker, who expects growth to occur in "fast and slow lane" applications. The fast category comprises such uses as rubber, plastics, and carbon fiber, where alterations to the manufacturing process aren't necessary and government approval isn't required. Water-purification membranes and biomedical applications will move more slowly, he explains, because of the need for regulatory approval and certification processes.

David Kramer

Submarines afford a view from below the Arctic

Civilian–military cooperation builds understanding of the ocean, Earth, and climate change and benefits national security.

Arctic scientists and the US Navy are breathing new life into a 27-year-old partnership, the Submarine Arctic Science Program (SCICEX). Its heyday was in the 1990s, when the navy hosted dedicated science cruises. Since then it has continued in a more hands-off and less scientifically productive mode: Although scientists receive submarine data from the navy, they are no longer able to embark, install their own specialized instruments, or determine expedition routes.

Those limitations will remain under a new memorandum of agreement that's in the works, says the University of Alaska Fairbanks's Jackie Richter-Menge, chair of the SCICEX science advisory committee. Still, she and other scientists anticipate new opportunities; in particular, they hope for quicker declassification of data collected by the navy's nuclear submarines.

BERNARD COAKLEY

THE USS *HAWKBILL* surfaced at the North Pole in 1999, during the Submarine Arctic Science Program's last dedicated science expedition. The *Sturgeon*-class attack submarine sailed from Honolulu, Hawaii, to Portsmouth, UK, in 67 days.

Security and science

SCICEX was the brainchild in the early 1980s of George Newton, a former navy captain who was a member, and later chair, of the US Arctic Research Commission, a government advisory body. (See the interview with current chair David Kennedy at PHYSICS TODAY online, 7 May 2021.) When the Cold War ended, says Newton, "we had built a submarine force that was in excess of the military's need in a peacetime environment."

Arctic submarine expeditions would both advance science and allow the navy to train sailors and maintain military capability in the harsh environment, Newton reasoned. "I started pushing the idea of the navy and scientists collaborating." The former Soviet Union also had an excess of submarines, he says. "They began offering submarines for lease to scientists. That served as a bit of a motivator for SCICEX."

By 2000, following a handful of collaborative expeditions, the partnership morphed. The navy had decommissioned most of its submarines that were suitable for Arctic missions, Newton says, and there were no longer enough of them for dedicated science cruises.

For the past two decades, the navy has tied collection of Arctic submarine data for scientists to the partly classified ice exercises (ICEX) it has set up at ice camps every year or two since the 1960s. It invites mainly navy scientists to do research at those ice camps.

"I want to reinvigorate the partnership," says Howard Reese, director of the navy's Arctic Submarine Laboratory in

San Diego, California, and the liaison to both ICEX science and SCICEX. "To really understand what's going on in the Arctic, scientists need to look from above and below."

Global warming has made surface passage across the Arctic easier, and for security purposes, the navy monitors activity in the area. It also has a long history of funding and running basic research in the Arctic. Submarine data help determine what's going on in terms of ice melt and climate change, says Reese. (See the articles by Peter Worcester and Megan Ballard, PHYSICS TODAY, December 2020, page 44, and by Martin Jeffries, James Overland, and Don Perovich, October 2013, page 35.)

Identifying patterns of freshwater runoff, temperatures, and sound velocity aid with national security, Reese says. For example, understanding physical features and the acoustic environment helps the navy to hide its submarines from potential adversaries and to listen for others' ships and submarines. And the navy needs to know where the ice is thin enough to surface a submarine. "We need to understand the environment to be able to operate effectively in the Arctic. Working with scientists is mutually beneficial," he says.

From seafloor to surface ice

Among the main quantities that scientists measure in the Arctic Ocean are seafloor topography; extent, thickness, and roughness of sea ice; and temperature, depth, conductivity, and nutrients. Some quantities can be measured via satellites, icebreaking ships, autonomous underwater vehicles, or instruments on the ice or in the water. But others are better—or only—accessible from a submarine.

Long-term changes in the thickness and roughness of sea ice are best studied from data collected with upward-looking sonar on submarines, says Richter-Menge. Satellites have done a good job of measuring sea-ice area since the 1980s, she says, but only recently have technological advances made it possible for them to infer thickness. Satellite estimates rely on models that take into account the properties of the sea ice and snow, she notes. "Contemporaneous submarine measurements offer the best tool for validating this new technology." Without submarine data, she says, "we wouldn't have appreciated the extensive thinning of the ice over the past few decades."

Geophysicist Margo Edwards, director of the University of Hawaii's Applied Research Laboratory, was the first woman to deploy on an operational navy submarine under ice. She remembers her second day out with SCICEX in 1999: "We were doing a survey of the Chukchi seafloor and we saw evidence of scouring—a sheet of ice had pushed around the terrain many thousands of years ago. It was exciting."

At that time, the SCICEX submarine was outfitted with an interferometric sonar device that provided topographical data of the seafloor in unprecedented extent and detail—at least an order of magnitude better resolution and more precise positioning than from standard surface single-beam echo sounders. The terrain data provide clues about the history of the planet and climate cycles, says Martin Jakobsson, a geophysicist at Stockholm University. He incorporates data from SCICEX into the Seabed 2030 Project, which aims to map Earth's entire ocean floor by the end of this decade. Many people use the project's data to model ocean circulation, study the seafloor, and more, he says.

Knowing the seafloor topography is necessary for navigating safely and laying pipes and cables, says Larry Mayer, director of the Center for Coastal and Ocean Mapping at the University of New Hampshire and a member of the SCICEX science advisory committee. It is important for defining countries' rights to resources, studying plate tectonics, predicting tsunamis, and locating shipwrecks to study maritime heritage, he continues. And the seabed topography influences ocean circulation. For climate models, "seafloor roughness is important in terms of generating turbulence that impacts the distribution of heat."

So far only about 20% of the global seafloor has been mapped to current standards. "Literally more of the Moon and Mars have been mapped, and at better resolution," says Mayer.

Bernard Coakley, a geophysicist at the University of Alaska Fairbanks, sailed on a pre-SCICEX test cruise in 1993 and then on later SCICEX expeditions. He and colleagues measured gravity anomalies in the Gakkel Ridge that "were difficult to explain unless it had a very thin crust." In 2001, subsequent dredging of the ridge, located between Greenland and Siberia, confirmed their gravity findings, he says.

Because of a history of irregular sampling and poor navigation, says Coakley, many Arctic geologic features "were mispositioned by 100 kilometers laterally and they might be 1500 meters shallower than had been thought." With submarines, he says, features of the water and seafloor can be sampled more systematically. Now that more is known about the Arctic, he says, "we can ask significant questions, like How did this feature form? and How do sets of faults on a ridge relate to each other?"

Ice camps

In the reconfigured SCICEX of the past two decades, scientists obtain some of the submarine data that are collected in conjunction with an ICEX ice camp. They also conduct experiments from the ice. The camps typically take place over six or seven weeks in late winter.

MIT emeritus professor Arthur Baggeroer and colleagues have studied a layer of warm water about 70 meters deep that enters the Arctic from the Pacific Ocean through the Bering Strait. Known as the Beaufort Lens, the warm water creates a barrier to sound penetration. The speed of sound is higher in the layer than in the surrounding, colder layers, so the layer acts as a refractive waveguide.

Discovered by Russian scientists, the layer "ensures the transarctic propagation of low-frequency sound," as described by Aleksandr Grigor'evich Litvak in *Herald of the Russian Academy of Sciences* (volume 85, page 239, 2015). Says Baggeroer, who spearheaded classified research on the water layer at ICEX, "I recognized the significance of the Beaufort Lens in the western Arctic for antisubmarine warfare."

Henrik Schmidt is director of MIT's Laboratory for Autonomous Marine Sensing Systems. At the 2016 ICEX camp he studied how the acoustic environment had changed. "We were lucky to study the same location that had been looked at in 1994. We compared directly the characteristics of ambient noise. There was a dramatic change." In 1994 the ice was 4 meters thick. Large ice floes would grind when blown by the wind, he says. In 2016, the ice buildup was 1 meter thick, and there was no more grinding. "If the wind blew, the noise was from ice cracking."

The biggest obstacle in working with ICEX, Schmidt says, is that "science is not the highest priority." Jon Collis, an underwater acoustician at MIT's Lincoln Laboratory, has taken part in the three most recent ICEX events. The navy plans to do tactical exercises during the 2022 ICEX, he says, "but they can't do that all the time, so it leaves a lot of time for science." Even so, competition is stiff for access to an ice hole, he says. "We negotiated and will work nights." His team plans to deploy sensors at various depths to measure salinity, temperature, and sound velocity and to use hydrophones to listen to the ice sheet breaking.

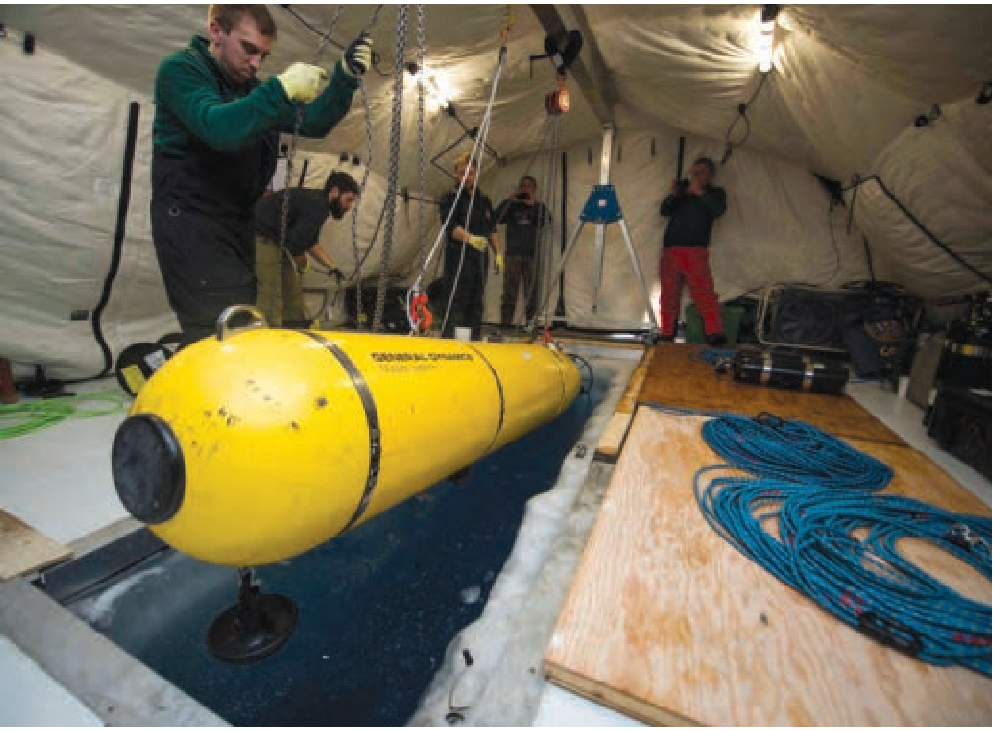
As for the current mode of SCICEX, without scientists on board the submarines, says Mayer, of the Center for Coastal and Ocean Mapping, "we have very little control of when instruments are on or off. We would, of course, rather use our own dedicated systems." Still, he says, the data that scientists obtain "are a lot better than nothing. Every sounding is useful." Scientists often can't afford submarines or Arctic camps on their own, so they're dependent on the navy for much of their Arctic access.

"The general perspective of the navy is that [hosting scientists] is a nice thing to do," says Val Schmidt, who was a junior naval officer aboard the 1998 and 1999 SCICEX cruises. "It's not the navy's primary mission, and they see it as a bit of a hassle having to accommodate the scientists." For his part, though, he says he had an "insatiable curiosity" and constantly "pestered" the scientists. After he left the navy he got a master's degree in ocean engineering and now leads the marine robotics program at the Center for Coastal and Ocean Mapping.

Partnering with the military

Military–civilian scientific partnerships to explore the oceans with submarines go back at least to the 1920s, says Sam

JOAN GARDNER (left) and Rick Hagen of the US Naval Research Laboratory analyze ice-core samples at the navy ice camp in March 2016. (Courtesy of the Arctic Submarine Laboratory.)



SCIENTISTS FROM MIT lower a sound array through a hole in the ice at the navy ice camp in March 2018. (Courtesy of the Arctic Submarine Laboratory.)

Robinson, a historian of ocean science at the University of Cambridge. At that time, Dutch geodesist and geophysicist Felix Andries Vening Meinesz took his instruments aboard Royal Netherlands Navy submarines in order to measure gravity anomalies. His goal was to establish the shape of Earth and the geoid, the shape Earth would take if winds and tides were absent.

For more than three decades starting in 1971, Peter Wadhams of the University of Cambridge sailed with UK Royal Navy submarines to study sea-ice thickness. Large ice blocks pile up to form deep ridges that protrude to depths of 40 meters or more. "That's important for navigation and for drilling rigs," says Wadhams. The distribution and size of the ridges follow a simple exponential law, he says. "You get fantastic insight into the role of ice and ice mechanics in climate."

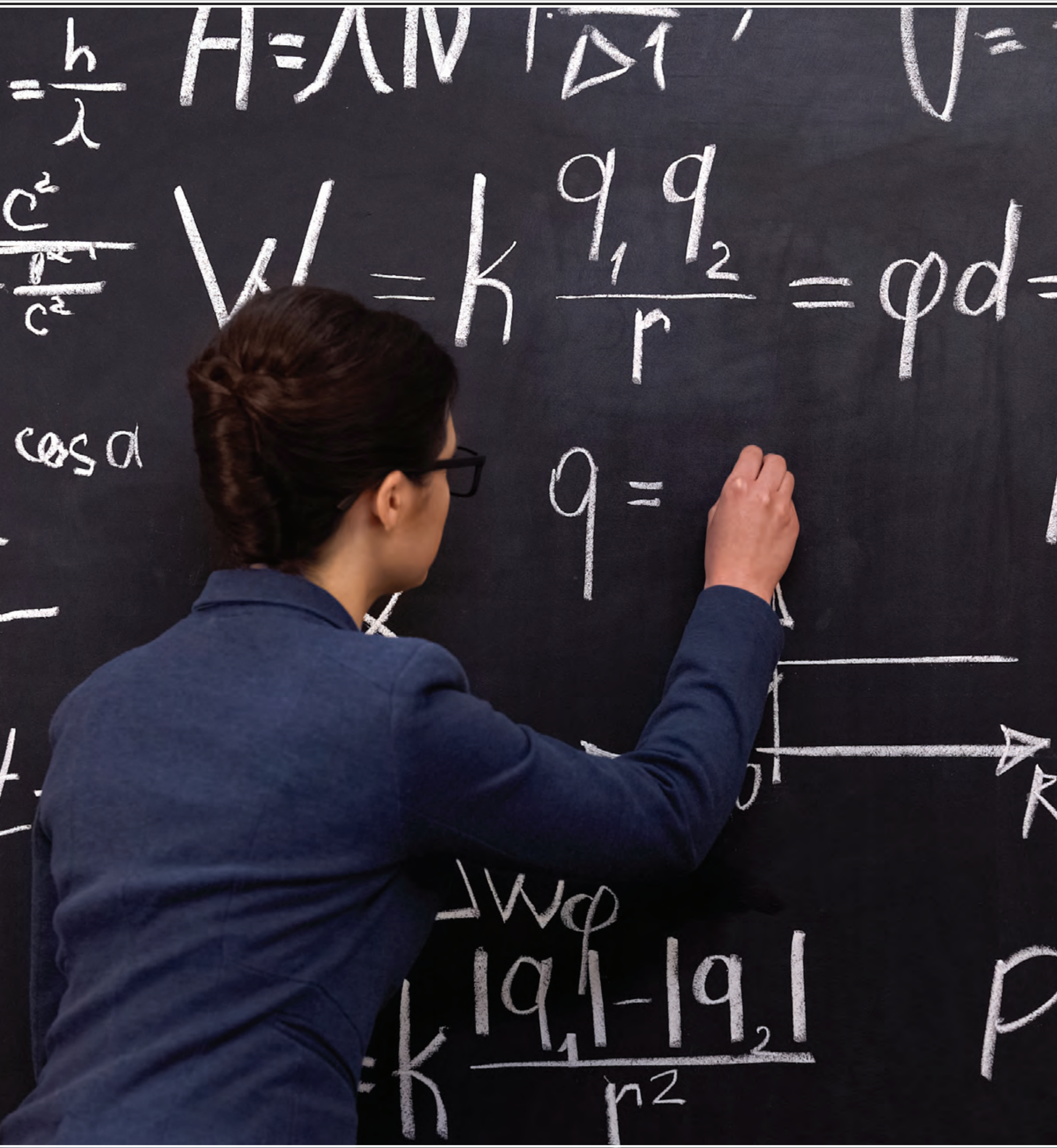
Wadhams's *in situ* explorations came to an abrupt halt in 2007 after a canister containing potassium chlorate that was part of the backup oxygen system exploded and killed two sailors. He now works with unmanned underwater vehicles. They are not as good as submarines, he says. "They are short range, and you can't collect data from across the entire Arctic basin."

A sticking point in all the military–civilian collaborations is that the data are classified. Even scientists who have security clearance and can access the data can't easily publish them in the open literature. The US Navy wants to keep the location and speed of its submarines under wraps: That's to prevent potential adversaries from identifying the acoustic signal of a given submarine by sifting through past recordings.

To that end, data released by the navy is "fuzzied up," says SCICEX science advisory committee chair Richter-Menge. More problematic is that, because the data are manually reviewed, it can take months or even years for the navy to release them to scientists. "The value of these data can depend on who's in charge," she says.

The navy has been "sporadic" about releasing data, says Reese, the Arctic Submarine Laboratory director. Algorithms are being developed to extract data that can safely be handed to civilian scientists. "If we get these data released, more scientists will be enthused and get involved," he says. "I hope we have a new agreement in a matter of months, not years."

Toni Feder PT


$= \frac{h}{\lambda}$
cos α
$q =$
$= \varphi d$
$k \frac{|q_1| \cdot |q_2|}{r^2}$

Mila Kryjevskaia is an associate professor of physics at North Dakota State University in Fargo. **Paula Heron** is a professor of physics at the University of Washington in Seattle. **Andrew Heckler** is a professor of physics at the Ohio State University in Columbus.



ISTOCK.COM/MOTORTION

Intuitive or rational? Students and experts need to be both

Mila Kryjevskaia, Paula R. L. Heron, and Andrew F. Heckler

Research into dual-process theories of reasoning from cognitive psychology suggests ways to improve classroom instruction in physics.

Suppose you are teaching an introductory-physics course. You ask your students to consider a box on a rough, level surface that remains at rest while a horizontal 30 N force is applied to it (see figure 1a). Your students seem to recognize that because the box is at rest, the frictional force must be equal in magnitude to the applied force. Using that realization, they correctly deduce that neither the mass of the box nor the coefficient of static friction is relevant for determining the magnitude of the frictional force.

As an instructor, you might be tempted to celebrate that little instructional victory. But then you ask your students a follow-up question about two identical boxes on surfaces with differing roughnesses, as shown in figure 1b. Both boxes remain at rest when pulled by horizontal forces of equal magnitude. You notice that many students seem to abandon the correct reasoning they formerly applied. Instead, they argue based on their intuition that the rougher surface under box C means that it must be subject to a greater frictional force.

The inconsistency in reasoning highlighted in that vignette may be familiar. Students who demonstrate desired knowledge and reasoning on one question often abandon that approach on closely related problems—even when those follow-up questions are posed just minutes or seconds later. That tendency can persist even after extensive instruction devoted to conceptual understanding and qualitative reasoning. As an instructor, you may wonder why students apply concepts, principles, and intuition in such a selective manner. More importantly, if students do possess the knowledge necessary to reason productively, why don't they catch their mistakes?

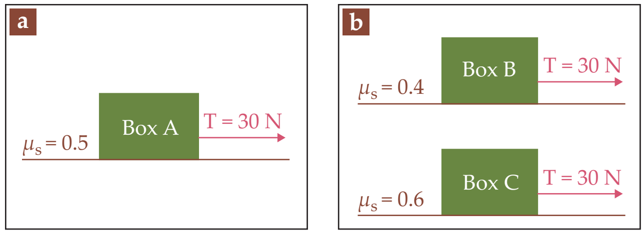


FIGURE 1. TWO PROBLEMS given to students in an introductory-physics class. **(a)** The initial problem involves a single box at rest on a rough, level surface while a horizontal 30 N force is applied to it. **(b)** The follow-up question involves two boxes at rest on surfaces of differing roughnesses; both boxes remain at rest when pulled by horizontal forces of equal magnitude.

Reasoning inconsistencies can be puzzling and frustrating to both instructors and students. Yet research in cognitive psychology suggests that they are a result of general human processes of thinking and decision making—and are thus normal, expected, and inevitable. In fact, studies show that experts use those same thinking processes to their advantage. Thus instructors should be able to anticipate and address inconsistencies in student thinking in a manner that encourages further learning and reduces frustration.

Our research on the teaching and learning of physics demonstrates that dual-process theories (DPTs) of reasoning, popularized by psychologist Daniel Kahneman, provide a useful lens for understanding how both novices and experts reason in physics.[1] DPTs model thinking in terms of both a fast, automatic process and a slow, effortful process. The theories allow us to pinpoint mechanisms contributing not only to incorrect answers and reasoning difficulties but also to productive expert-like reasoning pathways. In this article we describe how DPTs model human cognition and then apply them to the friction problem from the introductory vignette. We also present an analogy for DPTs that suggests some implications for physics instruction.

Dual-process theories

Since the early days of scientific inquiry into the human mind, researchers have worked toward developing general models of cognition that explain reasoning and decision making in all contexts. That work has led to DPTs, an influential family of models that are strongly supported by both neural imaging and laboratory studies that have examined participants' performance on various cognitive tasks.[2,3]

Most DPTs share the same key features. They model human cognition in terms of interactions between two processes: a fast, automatic, and subconscious process and a slow, effortful, and deliberate process. Following other researchers, we will refer to the former as process 1, or the "intuitive" process, and the latter as process 2, or the "analytic" process.

The differences between the two can be illustrated by the following question: "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?"[4] For most people, the answer 10¢ springs to mind from process 1. In fact, many immediately and subconsciously embrace that answer as correct without giving it any further consideration. Once it is revealed that it is incorrect, however, most can determine through process 2 that the correct answer is 5¢. Thus process 1 provides an immediate response; process 2 is needed to examine the situation more carefully and determine the correct one.

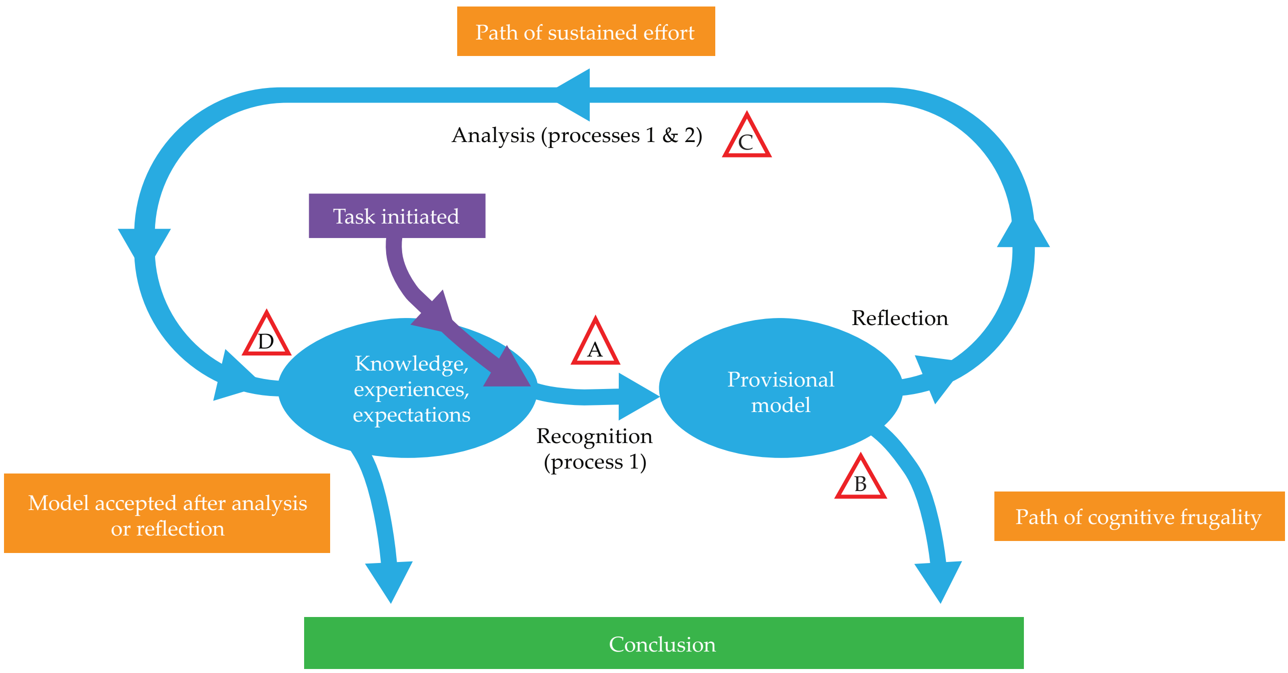
The interplay between a quick, automatic response and more deliberate processing can also be observed in student performance on physics problems, as illustrated in the opening vignette and other examples from our studies.[5–7] Specifically, when one is confronted with a new task, process 1 quickly makes associations between contextual cues and past experiences and produces a provisional mental model. In the case of our vignette, a student may associate more roughness with a greater frictional force and base their response on that model. The model cued by process 1 might be applicable in other contexts—that is why it is a learned association—but not necessarily to the situation at hand.

Incorporating intuition

DPT models also help researchers build a more formal characterization of intuition. Physicists value intuition and rely on it when practicing physics. Often, physics instructors stress that they want their students to develop physical intuition without carefully defining what that means. How can educators improve instruction toward a goal that is so vaguely defined? Although a precise definition of physical intuition remains elusive, DPTs offer some clues. Social scientist Herbert Simon once provided a parsimonious and pragmatic definition of intuition as "nothing more and nothing less than recognition."[8] In other words, intuition is the outcome of process 1: It is a quick and unconscious model or impression arising from associations based on past experiences. When a person says, "my intuition is that . . . ," they mean, "process 1 is telling me . . ."

Although physicists claim to value intuition, sometimes it is viewed negatively. For example, when a student's intuition leads to an erroneous response, their thinking is often described as everyday or informal and contrasted unfavorably with supposedly more sophisticated formal and analytical thinking. Yet intuition is also an indispensable component of expertise.[9,10] Experts possess an extensive repertoire of prior experiences, including recovery from errors, and learn to recognize appropriate cues and associate proper models. In fact, experts often automate formal processes so that they become effortless and intuitive. For instance, physicists may use intuition from their experience with conservation laws to constrain possible outcomes of a physical situation.

In the case of our introductory vignette, an expert will intuitively know that Newton's second law is the relevant concept and rapidly recognize that the question is about balanced forces. Their experience will lead them to correctly ignore seemingly alluring yet extraneous information, such as the differing coefficients of static friction. For the expert, intuition thus provides a quick, effortless solution and conserves cognitive resources for more challenging situations.



Possible hazards

A — Recognition, or intuition, provides an incorrect provisional model.

B — Reflection is cursory or nonexistent, so the provisional model is accepted without scrutiny.

C — Analysis is biased, flawed, or ineffective, so errors are not caught.

D — Knowledge is missing, weak, or misleading, so errors are not caught.

FIGURE 2. A SCHEMATIC MODEL of pathways—predicted by dual-process theories of reasoning—that students might take when attempting to answer a problem in physics. When a task or problem is initiated, students unconsciously recognize a provisional mental model based on their prior knowledge, experiences, and expectations. If they accept the provisional model, they proceed down the path of cognitive frugality and arrive at a quick conclusion that will likely be incorrect. If they instead choose to reflect on and analyze that provisional model, they take the path of sustained effort, which can be iterative. Eventually, students accept a model after analysis or reflection and arrive at a conclusion. There are four major points, indicated by the possible hazards A–D, at which errors in reasoning can occur.

Importantly, according to DPTs, the quick, subconscious process 1 cannot be turned off; we all perceive the world around us through its lens. It is overwhelmingly prevalent in everyday life, in which it is highly efficient and quite accurate in guiding judgments most of the time. As such, the output of process 1 serves as the entry point in most reasoning paths. The slow and deliberate process 2 may intervene to evaluate the output of process 1, but if a reasoner is confident in their first impression—for example, "the ball costs 10¢"—process 2 is circumvented and the first available mental model yields the final response. In physics, that can lead to reasoning pitfalls.

Reasoning pathways

To illustrate DPT models in greater detail, consider the opening vignette. In figure 2 we illustrate pathways to a solution to the problem, as predicted by DPTs of reasoning, along with four possible hazards, A–D.

The first step is recognition, which is led by process 1. After reading the question, the student subconsciously recognizes a provisional mental model based on prior experiences, expectations, and contextual cues. In the case of the vignette, the problem contains two main features that compete for attention. The salience of an irrelevant feature—namely, the differing coefficients of static friction μ_s—will distract many students from the relevant feature, which is that the boxes remain at rest. As a result, process 1 prompts many students to erroneously interpret the problem as being about frictional forces and surface roughness.

That intuitively appealing but incorrect model, illustrated by hazard A in figure 2, presents the first potential setback on the path to a correct conclusion. Other students—and presumably experts—recognize that the situation concerns an object at rest and requires an application of Newton's second law. That more accurate provisional model puts them on a path to a correct final response that avoids hazard A.

The second step involves a bifurcation: reflection versus final decision. After the initial mental model is formed, it can be scrutinized by the slow, logical, deliberate, and analytic process 2. However, if a reasoner feels confident in the accuracy of their first intuitive response, the analytic process may be circumvented, which results in a final decision based only on intuition. Such a direct path from the first intuitive mental model to a conclusion is said to stem from "cognitive miserliness," a universal human tendency to spend the least amount of effort on a cognitive task. We believe that student responses that contain nothing but a statement such as "friction is greater because μ_s is greater" are likely a product of that reasoning shortcut.

Direct paths from an intuitive mental model to a final response can sometimes be desirable. An expert whose provisional mental model is consistent with a normative response does not need to engage the analytic process 2 to answer correctly. For that reason, we prefer the term "cognitive frugality" to cognitive miserliness. Frugality on some tasks is necessary if more computationally demanding tasks are to be tackled. Knowing when it is safe to proceed with the output of process 1 and when it is important to scrutinize that output is essential. Reasoners with more highly developed cognitive reflection skills will navigate that juncture more effectively and engage the analytic process when appropriate.[4,11] Failure at that point is represented by hazard B in figure 2.

Analyzing mental models

If the reasoner avoids hazard B, the third step is explicit analysis. Here, again, there are several ways such an analysis can fail. For example, reasoners tend to struggle with searching for alternative solutions or generating counterarguments.[12] As such, the analytic process may be influenced by a process 1 form of confirmation bias in which cognitive resources are devoted to rationalization rather than error detection.

For example, a student who thinks that the coefficient of static friction μ_s is relevant to the problem posed in the opening vignette will find quick confirmation for that approach in a mathematical expression, $F_{fr} = \mu_s N$, where F_{fr} is the frictional force and N is the normal force that the surface exerts on the box. That expression is not applicable to the vignette, but it is often cited by students to validate a response they already believe to be correct. Such biased or flawed reasoning presents another significant obstacle on the path to a valid conclusion, and it is depicted by hazard C in figure 2.

Another threat to the validity of the analytic process is a lack of a sufficiently robust conceptual framework to evaluate the output of process 1. For example, some students who are successful at applying physics knowledge in straightforward situations, such as an object at rest, may not know how to use that knowledge to evaluate an intuitively appealing response in a more complex situation that involves significant distractions. They may rely on criteria other than consistency with Newton's laws to establish the validity of their argument. For instance, they might argue that "different coefficients of friction are given, so they must be relevant to the solution" or that "the formula $F_{fr} = \mu_s N$ has worked successfully in solving many other physics problems and therefore is likely to work here as well."

In fact, some tend to dismiss the outcome of qualitative reasoning if it appears to disagree with a mathematical formula because they perceive formulas to be higher in the hierarchy of reasoning tactics in physics. One student, for example, admitted to us that "people would probably think that the friction is the same in both cases because neither box moves, but after applying the formula for static friction, you can see the force of friction is different." In other words, an incorrect intuition-based mental model can often persist and lead students to an erroneous final response.

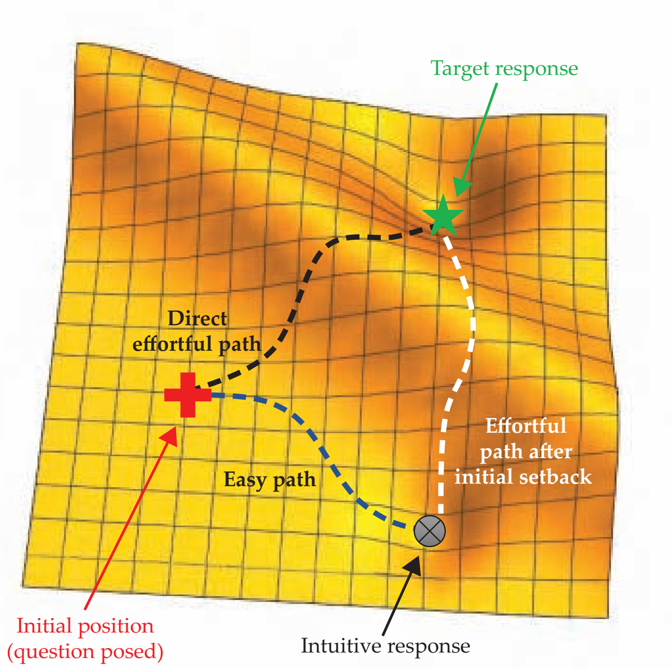


FIGURE 3. REASONING CHALLENGES in physics can be likened to a journey through treacherous terrain. The destination, or the target response, is the lowest point in the terrain. Students are likely to take the easy path to the intuitive response, represented by a depression that is easily mistaken for the destination. On the other hand, experts are likely to take the direct effortful path and scale the ridge with confidence. But even if experts are led astray to the intuitive response, their cognitive reflection skills may enable them to detect a navigation error and proceed to scale the ridge via the effortful path after initial setback.

If correct and relevant formal knowledge is absent or is not developed to the degree needed to support detecting and overriding errors, students who engage the analytic process 2 will likely be unsuccessful. That is indicated by hazard D in figure 2. One student described their failed attempt as follows: "I tried to formally think it through by trying to remember the magnitude equations and by thinking of the different forces or anything that may be relevant to the problem. I was uncertain with my formal thinking attempt, so I went with my intuitive reasoning."

Students who avoid hazard B—accepting the provisional mental model without scrutiny—are following a line of reasoning that we term the path of sustained effort, which is also shown in figure 2. That path can be iterative. If the first loop is completed without detecting red flags in reasoning, a response based on the provisional model is accepted and a conclusion is reached. However, if a reasoner identifies a need to consider alternatives, their process 1 may suggest a different provisional model cued by either the same or a different set of ideas, experiences, or expectations. In that case, the reasoning process persists and may undergo multiple iterations. As a result of that iterative reasoning process, the student may be able to successfully override the incorrect provisional model or models and reach a correct final response.

The more reasoning hazards a person must overcome during the analytic process, the more likely they will yield to the intuitive appeal of their initial model. The reasoning path that involves detecting and overriding errors is often cognitively expensive and emotionally unsettling. It is not surprising, then, that it is also frequently avoided. Adam, an introductory-physics student, provided a nearly perfect response to the question posed in the vignette: "[The] forces of friction are equal. . . . Since the boxes have equal force applied while remaining stationary, the friction will also be equal. The μ_s value is used more for maximum friction before being overcome." Adam not only compared the frictional forces correctly but also articulated why the coefficients of static friction are irrelevant to the question. Yet even he admitted some uncertainty upon finishing the friction task: "I applied the formal reasoning to the best of my ability, unless I thought myself into a wrong answer, which would look idiotic."[13]

The terrain analogy

Human reasoning's dual nature suggests that formal knowledge alone may not be sufficient for productive reasoning. Situations that tend to elicit intuitively appealing but incorrect responses are plentiful in physics. Because people cannot turn off the intuitive process 1, error detection mechanisms must be highly effective if they are to identify and override an initial mistake. Moreover, individuals must be motivated to tolerate the demands of the analytic process 2. We therefore argue that instructors should take the dual nature of reasoning into account and help students develop strategies for navigating reasoning challenges that can be likened to negotiating treacherous terrain.

The terrain analogy is useful for understanding instructional implications. Imagine that the task you set for your students lands them at an initial position in a hilly region (see figure 3). They don't have a map, but they expect their destination to be the lowest point in the terrain. Although they may not recognize it, the journey to the actual destination requires them to traverse a hill—or, in the problem, to expend some cognitive effort. However, process 1 provides them with an initial direction along an easy path. Arriving at a local minimum (the position marked as the intuitive response in figure 3), students may not recognize that it is not the actual desired destination.

Our opening vignette can be applied to the terrain analogy. The destination (realizing that the frictional forces are the same) is located on the other side of a steep hill that requires effort to traverse (overriding the appeal of the intuitive response and applying Newton's second law). The information about coefficients of static friction, however, suggests an easy stroll to an attractive depression in the terrain (larger μ_s means larger frictional force).

That location satisfies several criteria associated with the desired destination: an unambiguous answer supported both by prior experience with problems in which friction is linked to surface roughness and by mathematical authority in the form of a familiar equation. Because of those criteria, it is easily mistaken for the sought-after destination. Climbing out of the depression and exploring more terrain is rejected by many students as an unnecessary expenditure of effort (cognitive frugality).

Although some stay contentedly in that local minimum, expert reasoners quickly recognize that the easy path, although attractive, is not the way to their destination. Equipped with endurance, techniques, and superior travel gear—that is, motivation, reasoning skills, and formal knowledge—such experts are able to scale the ridge with confidence by using the direct effortful path to the target response, illustrated in figure 3. Even if experts are briefly led astray, their tendency to explore, using their cognitive reflection skills, may enable them to detect a navigation error, emerge from the alluring depression, and proceed to scale the ridge (indicated in figure 3 by the effortful path taken after the initial setback).

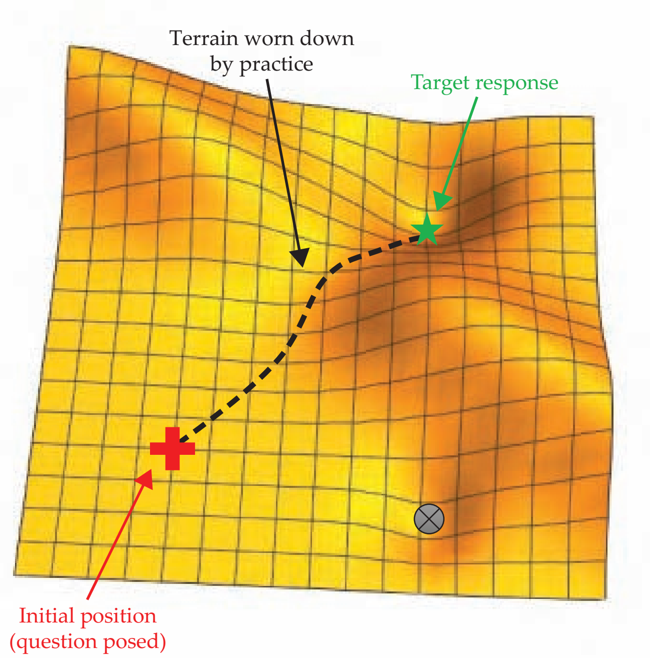


FIGURE 4. ANOTHER POTENTIAL PATH that experts may take through the treacherous terrain. Some of them have worn down a direct path through the mountain through years of extensive practice on similar problems.

Experts also have another potential scenario. Their path to the target may be so well traveled that it has altered the topography of the terrain itself, as shown in figure 4. Because those experts have worn down the path by years of extensive practice, they no longer require much effort to traverse it. For them, the automatic and effortless path suggested by process 1 and the desired path are now one and the same.

Instructional implications

The terrain analogy suggests two strategies for helping students. Both have benefits and drawbacks.

The first approach is to primarily focus on flattening the terrain and automating the direct path to the solution—that is, altering the likely outcome of process 1. If students repeatedly practice short sample problems, they will begin to recognize the path to the correct destination on similar problems. Such an approach will eventually wear down a direct path through the mountain range and allow students to determine the solution with little cognitive load. That means that process 1 is now likely to lead to a rapid, low effort and accurate outcome without engaging process 2.

One significant disadvantage to that approach is that it is unlikely to help students develop the skills necessary to navigate unfamiliar terrain and solve other kinds of problems. If they only receive instruction aimed at flattening the terrain, minor deviations from a well-traversed path or relatively small changes in the destination may leave students feeling lost, frustrated, and unprepared.

The second approach is to improve students' endurance, technique, and "hiking gear"—that is, to boost the likelihood that process 2 will be engaged successfully. It focuses on using deliberate reasoning, planning solution paths, analyzing alternatives, and evaluating results. To teach those higher-level skills, instructors should give students varied and challenging problems and help them thoughtfully and deliberately practice general reasoning and problem-solving strategies.

Part of that approach is aimed at helping students develop robust strategies for recognizing when they are on the right path or have reached the right destination—that is, detecting errors and overriding them. Research shows that quick fixes, such as using a single example to address novices' common intuitive responses, are unlikely to lead to a long-lasting effect.[5] But that does not mean that cognitive reflection cannot be taught. One effective way to help students develop general reasoning strategies is to elicit their intuitive process 1 responses to various problems and then provide guidance on how they can evaluate those responses.

Such an approach can help students recognize that some types of formal knowledge, such as Newton's second law, can be used both as a tool for solving a problem (for example, for an object at rest, forces must balance) and as a criterion for checking the validity and physical consistency of a provisional response (for example, if the static frictional force increases because μ_s increases, is Newton's second law still satisfied?).[6] In particular, we argue that instructors should help students cultivate the habit of questioning their intuitions. For example, students should be asking themselves such questions as the following: Is my model valid? What criteria must it satisfy to be valid? Is it consistent with other relevant physics principles? Returning to the terrain analogy, that approach should help students wander, get lost in a ravine, reflect, find a way out, and navigate through difficult terrain to the solution.

Training processes 1 and 2

There are trade-offs to the approaches described above. Although automating the solution path can optimize speed and reduce cognitive load, one cannot practically train on all possible problem paths. On the other hand, each task that arises should not require careful and effortful analytical thinking. Indeed, one hallmark of an expert is how they automate a significant number of tasks. Think, for example, about how effortlessly a physics instructor builds an integral.

We argue that some processes need to be automated for introductory physics, such as crucial math skills like trigonometric, algebraic, and vector operations. Improving students' fluency with essential skills enables their cognitive frugality on the basic steps needed for more complex reasoning. That, in turn, frees up their cognitive resources so they can engage in more computationally expensive analytic processes, such as reflection, the search for alternatives, and error detection. Those are particularly relevant for tasks for which an increased mastery of basic skills alone does not help students flatten the terrain sufficiently.

Overall, we propose that instruction should include both approaches. Focusing on process 1 is probably best for short, one- or two-step skills and knowledge. To improve such skills, teachers should use carefully designed, repeated, and spaced practice problems and give immediate feedback to build and retain fluency and reduce cognitive load.[14] In that way, students build one aspect of intuition by automating those skills. Initial studies on that type of approach are promising, although it remains an open question which skills lend themselves well to automation.

To help students develop approaches to productively engage process 2, instructors may need to provide repeated opportunities in multiple contexts. Such instruction will likely require making the interaction between the two processes visible to students. Instructors could present them with various scenarios, as seen in figure 2, in which hazard A is likely to be encountered. Those scenarios are likely to elicit misleading process 1 responses. Careful guidance could then be provided for navigating hazards B–D. As a result, students might learn how to reflect on an intuitively appealing response, use formal physics knowledge to check its validity, consider alternatives if a mistake is detected, and override a mistaken response.

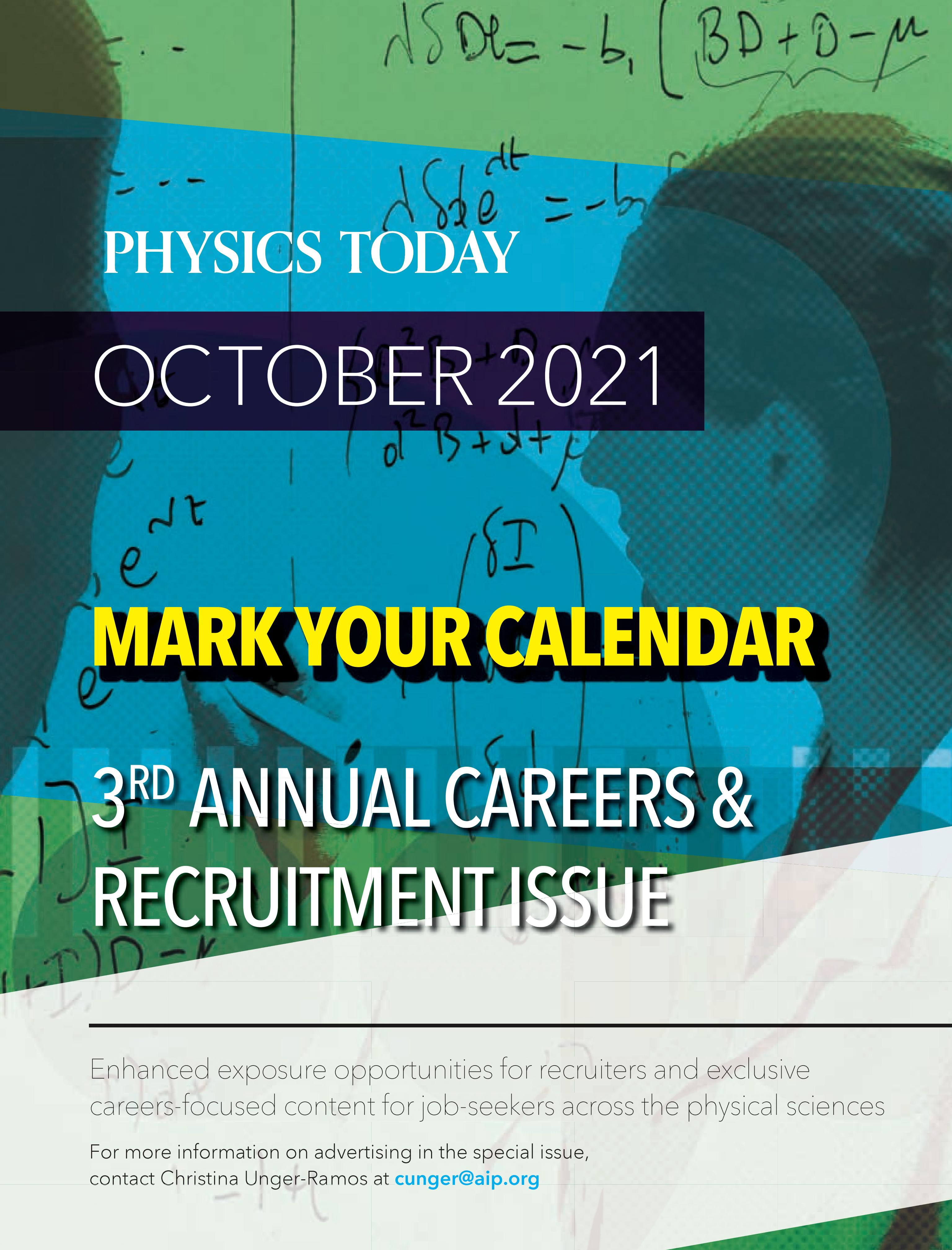
Finally, and perhaps most importantly, we speculate that a style of physics instruction that makes the dual nature of thinking explicit and visible to students may affect teaching in ways that extend beyond improvements in student performance. It may help establish an instructional environment that emphasizes that careful examination and possible rejection of an intuitive response is a natural part of reasoning and not an indication of a lack of knowledge or any other deficiency on the part of the student.

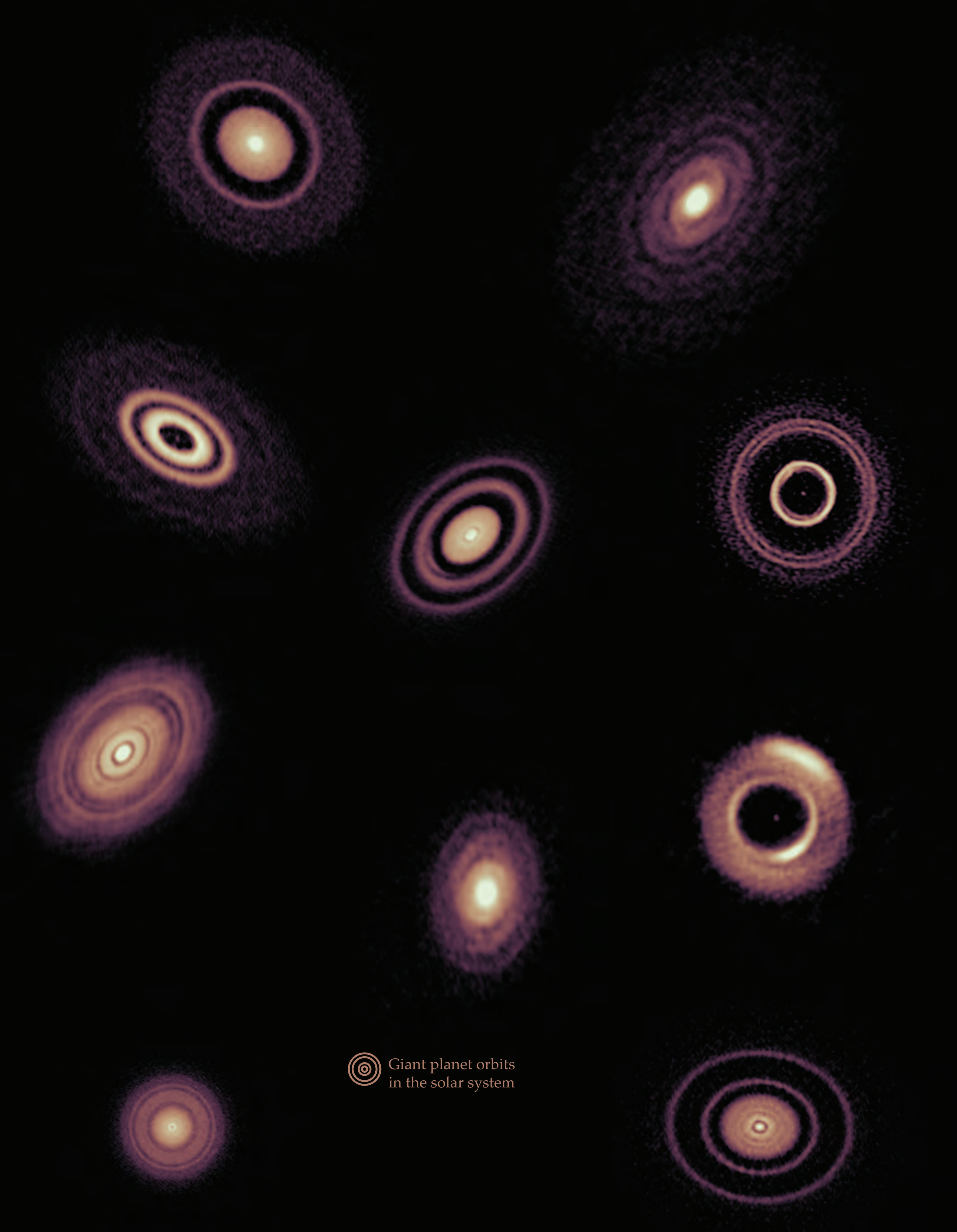
Rather than learn to reflexively discount their intuitions, students should be taught that intuition and formal knowledge are both important and can—and often do—interact fruitfully. Such a reframing may also help address a common concern that incorrect intuitive responses may lead to feelings of inadequacy, such as a student saying, "I am always wrong when it comes to physics, so I am not good enough to stay in the major." Classroom discussions emphasizing that initial incorrect responses often stem from the dual nature of human thinking may help alleviate such concerns and help bolster students' sense of self-worth and belonging in physics.

REFERENCES

1. D. Kahneman, *Thinking, Fast and Slow*, Farrar, Strauss and Giroux (2011).
2. D. Kahneman, *Am. Psychol.* **58**, 697 (2003).
3. J. S. B. T. Evans, *Annu. Rev. Psychol.* **59**, 255 (2008).
4. S. Frederick, *J. Econ. Perspect.* **19**, 25 (2005).
5. A. F. Heckler, T. M. Scaife, *Cogn. Sci.* **39**, 496 (2015).
6. M. Kryjevskaia et al., *Phys. Rev. Phys. Educ. Res.* **16**, 020140 (2020).
7. M. Kryjevskaia, in *Upgrading Physics Education to Meet the Needs of Society*, M. Pietrocola, ed., Springer (2019), p. 181.
8. H. A. Simon, *Psychol. Sci.* **3**, 150 (1992), p. 155.
9. C. Singh, *Am. J. Phys.* **70**, 1103 (2002).
10. L. Lising, A. Elby, *Am. J. Phys.* **73**, 372 (2005).
11. A. K. Wood, R. K. Galloway, J. Hardy, *Phys. Rev. Phys. Educ. Res.* **12**, 023101 (2016).
12. R. S. Nickerson, *Rev. Gen. Psychol.* **2**, 175 (1998).
13. Ref. 6, p. 18.
14. B. D. Mikula, A. F. Heckler, *Phys. Rev. Phys. Educ. Res.* **13**, 010122 (2017).

PT





Millimeter-wavelength continuum emission from protoplanetary disks, captured by the Atacama Large Millimeter/Submillimeter Array. (Data adapted from refs. 6, 7, 14, and 15.)

Sean Andrews is an astrophysicist on the staff of the Smithsonian Astrophysical Observatory, part of the Center for Astrophysics|Harvard & Smithsonian, in Cambridge, Massachusetts.

THE STRUCTURES OF PROTOPLANETARY DISKS

Sean M. Andrews

Astronomical observations of gas, dust, and rocky material in the disks from which planets emerge help refine theoretical ideas about how they form.

It is striking to recognize that we have been aware of planets orbiting other stars for only the past quarter century. More than 4300 of those exoplanets have been cataloged by astronomers in that time. Statistical extrapolations from that population suggest that more than one planet orbits each star, on average.[1] The big questions in astrophysics research have always been concerned with origins. The abundance of new worlds allows us to establish the context for our solar system and to refine ideas about the creation, evolution, and biological potential of planets. Of the many exciting discoveries made about exoplanets, perhaps the most profound is their diversity. No one knows how common a planetary configuration like our solar system may be, but it's clear that a much broader range of planetary properties exists than we can find among our immediate neighbors. Much of that variety is thought to be imprinted at the epoch of planet formation.

Planetary systems are assembled in the disks of gas and solid particles that orbit young stars. Those disks are byproducts of angular-momentum conservation during star formation, and they are created when a rotating condensation in a giant cloud collapses under its own gravity. The properties and evolution of the disk material determine where planets form, how they grow, what they are made of, and how their orbital configurations can change. In turn, feedback from that early development of a planetary system influences the behavior of the disk. The epoch of planet formation is brief: Observational signatures of disks disappear in less than 10 million years. But even the relatively short burst of mutual interactions with their disk birth sites profoundly influences the properties of the exoplanets and solar-system bodies measured today, billions of years later.

It is difficult to learn more about how those processes work when astrophysicists use only direct measurements of young planets. The common techniques for finding planets are less effective when the planets are in their youth, primarily because the enhanced magnetic activity of their stellar hosts can hide planetary signals. Instead, much of the effort and progress relies on astronomical observations of disks and *in situ* measurements of planets and other bodies in the solar system. The central goal of that work is to quantify the spatial distributions of the physical and chemical conditions in disks, particularly to hunt for the telltale signatures of dynamical interactions between the planets and disk material. Measurements of the disk properties for different evolutionary states and environments are essential for refining the theoretical ideas that link young planets in their formation epoch with their descendants—the populations of mature exoplanets and the planets in the solar system.

A planet-formation primer

Even early observers of the solar system had an implicit understanding that planet formation occurs in a rotating disk, because they recognized that the planets orbit the Sun in the same direction and confined to a narrow ecliptic plane. The development of a modern planet formation theory began in the 1960s and advanced along parallel tracks in the planetary science and astronomy communities. By the 1990s, those efforts had converged into the "core-accretion" paradigm. That model considers the evolution of a trace (approximately 1% by mass) population of planetesimals—solid building blocks, roughly comparable to large asteroids—embedded in a massive gas disk.

In the original theory, the swarm of planetesimals grows slowly, over roughly a million years, by constructive collisions into a population of rocky planetary cores. If the growth is efficient enough to reach a critical mass (a few times Earth's mass), then each core can rapidly acquire an atmosphere with tens to hundreds of times Earth's mass and form a giant planet like Saturn or Jupiter. The masses, initial orbits, compositions, and subsequent evolution of such planets are primarily controlled by the spatial distributions of density and temperature in their progenitor material: the disk.

One of the major challenges for the core-accretion model is associated with initial conditions: How does a disk produce planetesimals in the first place? The solids that a disk inherits from the interstellar medium are small—perhaps a few microns across. The model's viability hinges on the ability of those small particles to grow in size by at least 10 orders of magnitude within a million years. Numerical simulations of the particle growth, grounded in results from microgravity collision experiments, indicate that millimeter- to centimeter-sized pebbles are produced easily.[2] Subsequent growth is problematic, though. Classical models assume a smooth disk structure, in which gas densities and temperatures decrease monotonically with distance from the host star. That would be expected from gravitational-collapse models, in which the primary heating mechanism is stellar irradiation.

In that case, the dynamical contribution of pressure support means the gas orbits slightly below the Keplerian velocity. As solids in a particular size range decouple from the gas, the sub-Keplerian flow imparts a drag force that saps the solids' angular momentum and sends them spiraling in toward the global pressure maximum near the star. That process is especially effective for pebbles: They migrate faster than they can collide and grow in the planet-forming zones of most disks.[3]

One family of possible solutions to the planetesimal growth problem invokes two modifications to the simple picture of growth by binary collisions. First, and most importantly, the solutions assume that the gas distributions in disks must not be smooth. Local pressure maxima perturb the gas flow, and thereby the aerodynamics of the solids. Even relatively low-amplitude pressure deviations can slow or stop migrating pebbles.[4] Second, the solutions presume that the concentrations of solids near such pressure traps can become high enough to trigger a growth instability or direct gravitational collapse that rapidly converts pebbles into much larger solids.[5] The key unifying hypothesis is that the production of planetesimals—and therefore the viability of the core-accretion model of planet formation—requires that localized perturbations, known as substructures, in the physical conditions are robust and fundamental aspects of protoplanetary disks.

Astrophysicists have proposed many different physical origins for the hypothesized disk substructures. Most fall into two broad categories. In one class, perturbations are associated with various fluid-dynamics processes in the gas disk—from turbulence, to hydrodynamic instabilities, to coherent magnetohydrodynamical structures, and beyond. The basic idea is that the gas disk naturally generates its own substructures, which then act as pressure traps to concentrate migrating solids and facilitate planetesimal formation. Another class of proposals considers that disk substructures are generated by interactions with already-forming young planets. Once a planet is sufficiently massive, it can clear a gap in the material around its orbit, drive spiral arms that shock the gas well away from its location, seed hydrodynamic instabilities (vortices, for instance), and potentially perturb more of the disk through its own migration. The important distinction between the two categories is that the first is a cause of planet formation, whereas

FIGURE 1. ALMA. Shown here are a few of the 66 antennas that make up the Atacama Large Millimeter/Submillimeter Array interferometer, located on the Chajnantor Plateau in the Atacama Desert in Chile. Over the past few years, ALMA's exquisite sensitivity and resolution have revealed that the disks around young stars are riddled with substructures. The details are revolutionizing our understanding of the planet formation process. (Image by D. Korden/ESO.)

the second is an effect of it. They are not mutually exclusive, but the latter implicitly presumes that an earlier generation of disk substructures was created by the former.

The ALMA revolution

Measurements of the prevalence, locations, morphologies, sizes, and amplitudes of disk substructures are essential for assessing the roles of those mechanisms. Put simply, a deeper understanding of planet formation requires observations that find and characterize substructures. Pebbles are the optimal probes. They should show the highest-amplitude perturbations in their spatial distributions, given their concentrations near gas pressure maxima. Pebbles absorb and re-emit starlight as thermal continuum radiation, with a peak efficiency at the millimeter wavelengths detectable by radio telescopes. Although contrast is highest for the millimeter-wavelength intensity variations from substructures, resolution is a challenge. Stable perturbations to the gas disk have a characteristic size comparable to the pressure-scale height H (the ratio of the local sound speed and the Keplerian angular velocity). For typical disk properties, H is 5–10% of the roughly 10–100 AU separation from the host star. That size implies that most substructures span only a few astronomical units. (An AU, the mean Earth–Sun distance, spans roughly 1.5×10^{11} m.)

Given the distances to the nearest disks, those expected sizes suggest an angular-resolution requirement of 0.03 arcsec. That's equivalent to resolving an airplane on the Moon. At a wavelength of 1 mm, that resolution corresponds to the diffraction limit for a telescope with a diameter of 10 km. Because a single aperture that size is an engineering impossibility, the solution is to link together many smaller apertures and form an interferometer. The necessary combination of sensitivity and resolution only recently came together with the commissioning of the revolutionary Atacama Large Millimeter/Submillimeter Array (ALMA), shown in figure 1. Constructed by an international partnership of 22 countries, ALMA is the largest ground-based astronomical observatory in the world. Located in the Atacama Desert in Chile, it is designed to measure continuum and spectral line emission at wavelengths of 0.3–7 mm with exquisite sensitivity. The 66 antennas can be arranged to probe spatial scales comparable to a telescope with a 16 km diameter.

Even before the highest-resolution configurations of the ALMA antennas were commissioned, hints of disk substructures were emerging in images that probed scales down to 10–20 AU. They included observations of "transition" disks, a subgroup representing about 10% of the population of all disks and defined by the depletion of continuum emission in a large inner zone, tens of astronomical units from the host star. But notably, annular variations of the continuum emission were being serendipitously uncovered in a few disks that did not fall into that category.

In the fall of 2014, the ALMA staff first experimented with the most extended antenna configurations to achieve very high resolution. One of their targets was the disk around the very young star HL Tauri, previously argued to have a smooth emission distribution on 15 AU scales. The new ALMA measurements at 4 AU resolution showed a stunning morphology riddled with substructures—narrow, concentric, dark gaps and bright rings distributed to about 100 AU from the star (see the first disk[6] above the one in the bottom left in the gallery on page 36). In the next available observing season that year, I worked with a team that used a similar ALMA antenna configuration to measure another ring and gap morphology in the disk around TW Hydrae (see the bottom left image[7] on page 36). As a few more cases followed, it became clear that the goals of finding and characterizing disk substructures were readily achievable with ALMA.

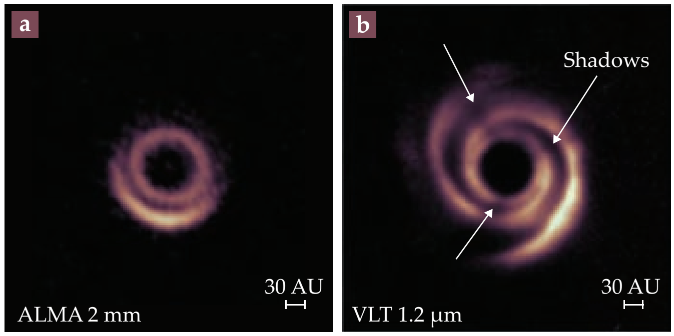


FIGURE 2. THE DISK around the young star SAO 206462, viewed in **(a)** the thermal continuum emission at a wavelength of 2 mm, from the Atacama Large Millimeter/Submillimeter Array (ALMA), and **(b)** the scattered starlight at a wavelength of 1.2 μm, from the 8.2-m-diameter Very Large Telescope (VLT). The strikingly different morphologies indicate the decoupled behaviors of the large particle in the midplane (panel a) and small solid particles in the atmosphere layers (panel b) of the disk. The narrow shadows (marked by arrows) in the scattered-light image suggest vertical perturbations at smaller separations from the central host star, not seen here. (Panel a adapted from ref. 16; panel b adapted from ref. 17.)

Over the past few years, high-resolution ALMA observations have continued to uncover examples.[8] Substructures are ubiquitous—they are found in all disks that have been explored with sufficient resolution to identify features at sizes comparable to the local scale height. The most common morphology is a set of narrow, concentric, and symmetric rings and gaps; a subset of cases shows more complex spirals or arc features. In the arc morphologies, rings and gaps are often also present. Substructures do not appear to have any preferential locations: They are found at any separation from the host star, from the inner resolution limit of a few astronomical units to the outer reaches where the millimeter continuum can be detected, more than 200 AU in some cases.

Complementary insights

Although researchers have learned much about disk substructures in a short time, the samples available to study are still relatively few and biased. The focus so far has been on the brighter, larger disks that are preferentially found around stars with masses comparable to the Sun or larger. Although the samples of millimeter continuum measurements help probe systems more representative of the general disk population, the immediate focus in the field has shifted to more-detailed explorations of substructure properties in individual case studies. An important aspect of that work involves a joint analysis with complementary tracers of disk material, particularly those that are more directly sensitive to substructures in the gas phase.

One such tracer relies on the optical and IR starlight reflected by micron-sized dust particles suspended in the disk

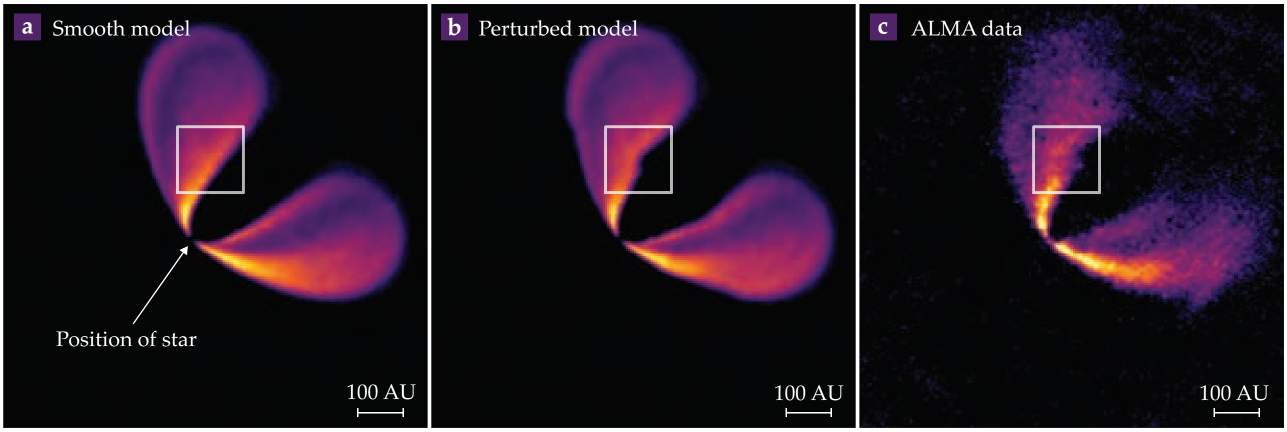


FIGURE 3. **KINEMATIC SUBSTRUCTURES** reveal themselves in subtle deviations from a smooth, Keplerian rotation model in spatially and spectrally resolved maps of molecular spectral line emission. These panels show the carbon monoxide emission maps for **(a)** a smooth model, **(b)** the same model perturbed by a planet twice as massive as Jupiter orbiting at 260 AU, and **(c)** real Atacama Large Millimeter/Submillimeter Array (ALMA) observations of the disk around a young star. The kinks in the white boxes of panels b and c are indicative of a kinematic perturbation to the gas flow. Note that the panels plot only a portion of the spectral line and don't resemble the usual morphology of a disk. (Models in panels a and b courtesy of Jaehan Bae; data in panel c adapted from ref. 14.)

atmosphere—the gas that extends vertically above the dense disk midplane.[9] With the *Hubble Space Telescope* or large ground-based observatories with adaptive optics systems to correct for the blurring effects of Earth's atmosphere, the scattered light can be measured at a resolution comparable to ALMA's. In many cases, the observed morphology is quite different from the millimeter continuum emission. Figure 2 shows a striking example. Various factors contribute to those contrasting appearances but the key insights are that each tracer probes a different vertical location in the disk and a different phase of the disk material (either the small dust grains that are coupled with the gas or the pebbles that are not). A combined analysis of those complementary tracers can be used to study the three-dimensional behavior of substructures and the associated gas–solid interactions, both of which are crucial for discriminating between models of substructure origins.

Another interesting aspect of scattered-light images is their sensitivity to vertical substructures in the inner disk, even well below the resolution limit. Vertical perturbations at small disk radii can block starlight from illuminating the material at much larger distances. The intensity variations induced by those shadows are common, and sometimes their motions can even be tracked over time. That behavior is often attributed to a warped geometry, an orbiting perturber, or stochastic upwelling of material near the inner disk edge. In any case, the shadows suggest again that substructures are 3D and persist even on very small scales.

A complementary option for tracing substructures is to directly probe the gas reservoir. Most of the disk mass is cool molecular hydrogen gas, but because of its lack of a permanent dipole moment and typical disk conditions, that reservoir of material is effectively dark. ALMA, however, can observe the spectral line emission from pure rotational transitions of other simple molecules, such as carbon monoxide, to probe the gas distribution.[10] Those measurements can be made simultaneously with the millimeter continuum and therefore achieve comparable resolution. The emission from different spectral lines probes the gas in different vertical layers. Combined measurements of the lines allow researchers to reconstruct the 3D behavior of substructures based on the variations in their line intensities and ratios. Moreover, comparing the behavior of various gas species can help disentangle any compositional changes associated with those substructures. That mapping of spectral-line-intensity variations to reveal more about substructures is really just getting started. Even so, early results are promising.

Alternatively, the same spectral line data can be used to probe substructures through their kinematic perturbations of the gas velocities.[11] ALMA observations provide spatially resolved maps of the spectral line emission at resolutions of 100 m/s or better in projected velocity. That's sufficient to measure deviations as small as a few percent from a smooth Keplerian flow. The larger-amplitude deviations can be seen directly in spectral line images, as demonstrated in figure 3. Observations of those deviations in multiple spectral lines could be used to reconstruct the 3D behavior of the gas flows near substructures and thereby determine the associated pressure gradients around the local maximum. Linking the dynamical information with the complementary, intensity-based constraints from spectral lines, the millimeter continuum, and scattered light promises to provide detailed quantitative constraints on the substructure properties. That information, coupled with hydrodynamics simulations, will be essential for determining the mechanisms responsible for generating the foundational perturbations in disks.

Current landscape, new directions

Planet formation research has experienced a profound shift in the past few years, as observations have forced astronomers to appreciate that small-scale substructures are an essential aspect of the disks around young stars. As I've mentioned, those perturbations to the conditions of a disk appear to be ubiquitous and are usually symmetric and small, with sizes comparable to the local pressure-scale height. The available samples show that they do not seem to have a preferential location in the disk or any association with their host stars. Concentric

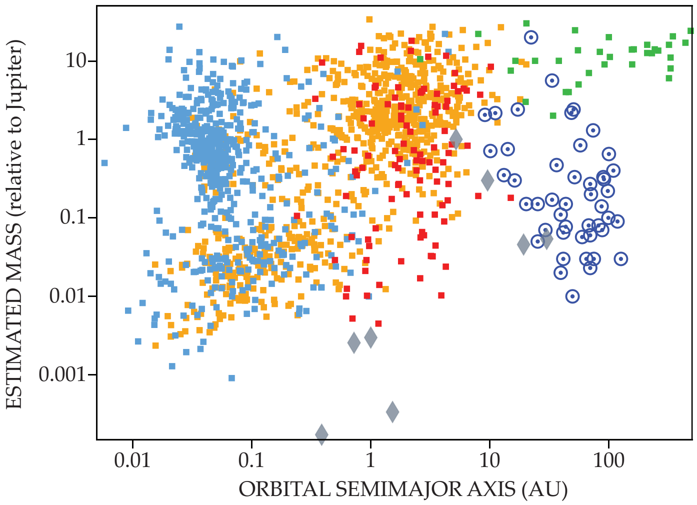


FIGURE 4. **ESTIMATED MASSES** and orbital locations for hypothesized planets (blue) that would clear the gaps in young disks observed with the Atacama Large Millimeter/Submillimeter Array,[12,18] compared with the known population of exoplanets (squares) and planets (diamonds) in our solar system. The exoplanet data are colored according to how they were detected: light blue for transits across a star, orange for radial velocity variations, red for gravitational microlensing, and green for direct imaging. (Data collated from the NASA Exoplanet Archive.) The masses and orbital locations inferred from disk substructures mostly occupy a swath of parameter space that has not yet been well searched for mature exoplanets. New facilities, however, will provide access to the older exoplanets with those masses and orbital locations and thereby enable tests of planetary systems' evolutionary pathways.

rings and gaps seem to be the most common morphology. Although researchers have learned much about substructures, they are still in the early stages of interpreting high-resolution data to determine the substructures' properties.

In general, the mechanisms that form those substructures remain unclear. Researchers do not yet know whether they are sites of planetesimal formation or signs that planetary systems already exist and are interacting with material in their birth sites. For some of the most striking examples, support is growing for the latter possibility. Much of the confidence in that hypothesis comes from hydrodynamic simulations of planet–disk interactions, which predict the gaps and rings observed in the millimeter continuum emission, and from the few cases that show pronounced and spatially localized perturbations in the spectral line emission.[12] Moreover, at least one case exists—the remarkable PDS 70 system—in which young giant planets still accreting their own circumplanetary material have been directly imaged and are clearly associated with disk substructures.[13]

If support for that possibility firms up and can be generalized, the simulations suggest that giant planet formation is well underway at 10–100 AU separations at ages less than a million years old. That's much faster than would naively be expected in the classical core-accretion model. Figure 4 shows the best estimates of young planet masses and orbits inferred from disk substructures, relative to the known planet population. The efficiency implied by those inferred planetary properties suggests that the epochs of planet and star formation overlap. Planetesimal and planetary core formation are therefore more robust than one would traditionally expect. An important test of that idea is to search for substructures in even younger disks, still embedded in their star-forming envelopes.

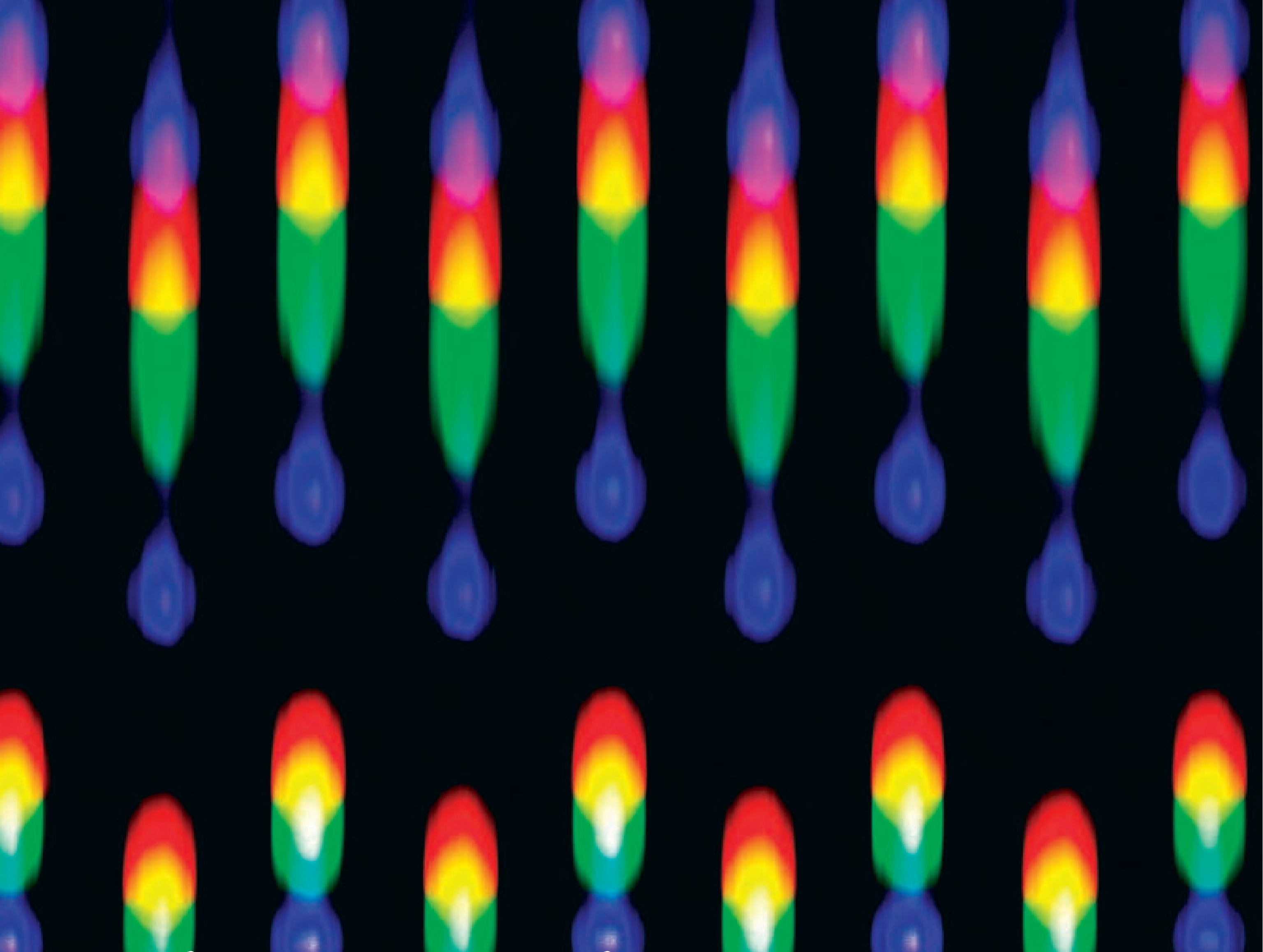
Ultimately, the final assessment of the viability of the planet–disk interactions hypothesis will come through direct imaging searches for young planets in the gaps in those disks. Some of that work is already underway and takes advantage of the fact that those systems shine substantially brighter than just their photospheres, because of their own compact circumplanetary dust disks and associated accretion flows. But the near future promises a considerable improvement in the sensitivity to lower-mass planets at smaller separations from their host stars. That improvement will come from the *James Webb Space Telescope*, planned to launch this autumn, and a new suite of large, 30-m-diameter ground-based observatories planned to be constructed over the next decade. Eventually, statistical comparisons of the distributions in mass and orbital semimajor axes for the young planets and the older exoplanet populations will provide a definitive constraint on planetary migration histories.

Studying disk substructures in increasing detail is also exciting. Astronomers are looking forward to improvements in existing facilities, such as ALMA, and to the development of new observatories that will push the resolution frontier. For example, the proposed next-generation Very Large Array, a radio interferometer that will specialize in longer, centimeter-scale wavelengths, will be able to resolve substructures at sub-astronomical-unit scales in the innermost parts of nearby protoplanetary disks. The combined access to such features from new and existing facilities promises that protoplanetary disk observations will continue to help refine our understanding of planet formation in the years to come.

I thank David Wilner and Rich Teague for their comments and suggestions, and Jaehan Bae, Myriam Benisty, and John Carpenter for kindly sharing their results to help with the figures.

REFERENCES

1. J. N. Winn, D. C. Fabrycky, *Annu. Rev. Astron. Astrophys.* **53**, 409 (2015).
2. T. Birnstiel, M. Fang, A. Johansen, *Space Sci. Rev.* **205**, 41 (2016).
3. S. J. Weidenschilling, *Mon. Not. R. Astron. Soc.* **180**, 57 (1977).
4. P. Pinilla et al., *Astron. Astrophys*. **538**, A114 (2012).
5. E. Chiang, A. N. Youdin, *Annu. Rev. Earth Planet. Sci.* **38**, 493 (2010).
6. ALMA Partnership et al., *Astrophys. J. Lett*. **808**, L3 (2015).
7. S. M. Andrews et al., *Astrophys. J. Lett*. **820**, L40 (2016).
8. S. M. Andrews, *Annu. Rev. Astron. Astrophys*. **58**, 483 (2020).
9. A. Garufi et al., *Astron. Astrophys*. **620**, A94 (2018).
10. A. Isella et al., *Phys. Rev. Lett*. **117**, 251101 (2016).
11. C. Pinte et al., *Astrophys. J. Lett*. **860**, L13 (2018).
12. S. Zhang et al., *Astrophys. J. Lett*. **869**, L47 (2018); J. Bae, Z. Zhu, L. Hartmann, *Astrophys. J*. **850**, 201 (2017).
13. M. Keppler et al., *Astron. Astrophys*. **617**, A44 (2018); S. Y. Haffert et al., *Nat. Astron*. **3**, 749 (2019).
14. S. M. Andrews et al., *Astrophys. J. Lett*. **869**, L41 (2018).
15. J. Huang et al., *Astrophys. J*. **891**, 48 (2020); S. Pérez et al., *Astron. J*. **158**, 15 (2019); R. Dong et al., *Astrophys. J*. **860**, 124 (2018).
16. P. Cazzoletti et al., *Astron. Astrophys*. **619**, A161 (2018).
17. T. Stolker et al., *Astrophys. J*. **849**, 143 (2017).
18. J. Bae, P. Pinilla, T. Birnstiel, *Astrophys. J. Lett.* **864**, L26 (2018); G. Lodato et al., *Mon Not. R. Astron. Soc.* **486**, 453 (2019). PT



MICROCHIP ACCELERATORS

R. Joel England, Peter Hommelhoff, and Robert L. Byer

An international collaboration aims to couple ultrafast lasers with integrated photonics to create chip-scale devices.

AVKOST/SHUTTERSTOCK.COM

Joel England is a lead scientist and head of accelerator operations for the Ultrafast Electron Diffraction Group at SLAC National Accelerator Laboratory in Menlo Park, California. **Peter Hommelhoff** holds the chair of laser physics and is a professor of physics at Friedrich–Alexander University Erlangen–Nuremberg in Germany. **Robert L. Byer** is the William R. Kenan Jr Professor of the School of Humanities and Sciences, a professor of applied physics at Stanford University, and member of the SLAC photon science department.



Particle accelerators are among the most important scientific tools of the modern age. Major accelerator facilities, such as the 27-km-circumference Large Hadron Collider in Switzerland, where the Higgs boson was recently discovered, allow scientists to uncover fundamental properties of matter and energy. But the particle energies needed to explore new regimes of physics have increased to the TeV scale and beyond, and accelerator facilities based on conventional technologies are becoming prohibitively large and costly. Even lower-energy, smaller-scale accelerators used in medicine and industry are often cumbersome devices; they can weigh several tons and cost millions of dollars.

Efforts are consequently underway to develop more compact, less expensive accelerator technologies. One approach, a dielectric laser accelerator (DLA), uses an ultrafast IR laser to deliver energy to electrons inside a microchip-scale device. Efficient, ultrafast solid-state lasers and semiconductor fabrication methods developed over the past two decades have enabled a new breed of photonic devices that can sustain accelerating fields one to two orders of magnitude larger than conventional microwave-cavity accelerators.

The approach has the potential to dramatically shrink particle accelerators, thereby enabling ultrafast tabletop electron diffraction and microscopy experiments and tunable x-ray sources. An international effort is now underway to develop a laser-driven accelerator integrated on a silicon photonics platform: an "accelerator on a chip."

Technological evolution

A conventional particle accelerator uses a series of metallic cavities powered by microwave energy to continuously accelerate bunches of charged particles traveling through it. Similar to a surfer riding on an ocean wave, particles in the accelerator ride an electromagnetic wave and either gain or lose energy depending on whether they are located at a peak or trough along the radiation's wavelength, which is on the order of 10 cm in conventional accelerators.

Shortly after the first laser was demonstrated at Hughes Research Laboratories in 1960, scientists began envisioning ways to harness the newly realized power for particle acceleration. The idea was appealing because the corresponding reduction in wavelength by four orders of magnitude—from the RF regime of conventional microwave accelerators to the optical regime of lasers—implied shrinking accelerator structures down to the micron scale.

But the required technology for laser-driven accelerators did not exist in the 1960s. It took several decades for solid-state lasers, optical materials, and nanofabrication to reach the point where researchers could pursue the idea in earnest. Also, metals were the prime candidates for shaping the accelerating fields, but they can't withstand the necessary large optical power levels because they absorb too much. Then, in the 1990s, researchers had the idea to use optically transparent dielectrics. That idea enabled the creation of the Laser Electron Acceleration Program (LEAP), which was based at Stanford University and SLAC in the early 2000s and aimed to realize a fully laser-driven particle accelerator.

LEAP led to one of the first demonstrations of laser-driven acceleration. The scheme used the so-called inverse transition radiation effect, in which a laser was shone onto a material boundary—a thin gold foil—that sat in the path of a high-energy electron beam.[1] The laser's electric field terminated at the boundary on a half cycle, so the truncated laser field gave the particles an accelerating "kick." The interaction distance in the proof-of-principle experiment, however, was limited to one half of an optical cycle, or less than 1 μm.

A key realization from those early experiments was that to make an efficient laser-driven structure, one should take advantage of the high damage threshold and low losses of dielectric and semiconductor materials. The laser in the LEAP experiments had to operate above the damage threshold of the metal surface to produce a measurable electron energy gain, so the gold film was mounted on a reel that refreshed the surface between each laser shot. Transparent dielectrics and semiconductors, on

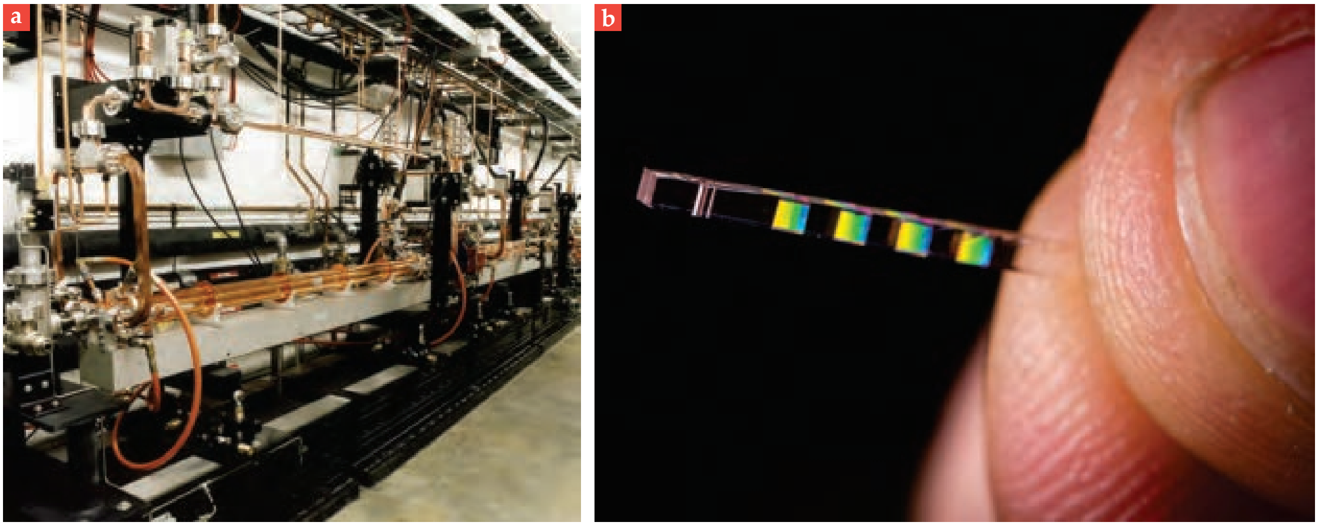


FIGURE 1. THE SIZE CONTRAST between conventional accelerator facilities and chip-based accelerators is dramatic. **(a)** The Next Linear Collider Test Accelerator facility at SLAC was used for early laser-acceleration experiments in 2012–15. (Image courtesy of the Archives and History Office/SLAC National Accelerator Laboratory.) **(b)** The first dielectric laser accelerator chips demonstrated at SLAC were made of fused silica and were each the size of a grain of rice. (Image courtesy of Christopher Smith/SLAC National Accelerator Laboratory.)

the other hand, can sustain accelerating field strengths that greatly exceed those of conventional particle accelerators. Various all-dielectric high-gradient accelerator designs have since been proposed. (For a review of DLA designs, see reference 2.)

In 2011 the authors of this article forged a collaboration with Purdue University and Tech-X, a commercial accelerator code developer in Boulder, Colorado. The program, AXiS (Advanced X Ray Integrated Sources), was a successor of LEAP and was sponsored by the US Defense Advanced Research Projects Agency. It was aimed at developing ultracompact accelerator technologies for portable medical imaging applications. Under the newly formed collaboration, the first high-gradient demonstrations of laser acceleration in nanostructured devices were conducted in parallel experiments at SLAC and at Friedrich–Alexander University (FAU) Erlangen–Nuremberg in Germany in collaboration with the Max Planck Institute of Quantum Optics (MPQ).

A tale of two experiments

Both the SLAC and FAU/MPQ experiments were conducted with fused silica devices. To make the tiny chips work as particle accelerators, periodic features were precisely fabricated along a vacuum channel through which the electrons would propagate. The periodic ridges spatially modulated the incoming laser phase fronts to match the accelerated electrons. Because the wavelength of the near-IR light that generated the force on the electrons was close to 1 μm, the dimensions of the internal structure's features and electron channel had to be around the same size.

A good first step for demonstrating a new acceleration mechanism is to send a test beam of particles through the device and observe the particles' energy distribution, or spectrum, following the interaction. The SLAC experiment used the 60 MeV electron beam provided by a conventional RF accelerator, the Next Linear Collider Test Accelerator, that had been converted into a test bed for developing new concepts. In a David-and-Goliath juxtaposition of size scales, the first DLA trial used a 50-m-long conventional accelerator facility to test some of the smallest particle accelerators ever built (see figure 1).

In late February 2013, two years of preparation finally paid off. As an automated shutter turned a laser on and off, the energy spectrum of the electron beam passing through the device shifted dramatically and unmistakably, thereby indicating the presence of large accelerating fields. Stanford graduate students Edgar Peralta and Ken Soong volunteered to run the

FIGURE 2. THE CONTROL ROOM at SLAC's Next Linear Collider Test Accelerator facility was the setting for an overnight experimental run led by Stanford students Edgar Peralta (right) and Ken Soong. Their successful experiment was one of a pair that would change the trajectory of research in dielectric laser acceleration and set the stage for an international effort to develop miniaturized particle accelerators. (Image courtesy of Ken Soong.)

experiment overnight to collect as much data as possible before the experiment shut down. At 4:05am they took the photograph shown in figure 2. Both were smiling because they knew they had reached an important milestone—and had likely secured enough data to complete their PhDs.

Subsequent analysis of the data showed that the measured variation in the final electron energy corresponded to an accelerating gradient, or energy gain per unit length, in excess of 300 MeV/m, which is about 10 times as high as typical operating gradients in conventional microwave accelerators. A scanning electron micrograph image of the DLA structure and corresponding field pattern are shown in figure 3a.

Around the same time that the experiment at SLAC was taking place, graduate student John Breuer, working with one of us (Hommelhoff) at MPQ, performed a similar experiment. It used a smaller and lower-energy (28 keV) electron source based on a modified electron microscope column. The column came from an old, malfunctioning electron microscope at MPQ and required three years of refurbishment to become an electron source for DLA experiments.

The accelerator structure consisted of an optical grating that was open on one side, as shown in figure 3b. In general, enclosed structures offer better field uniformity and improved coupling of the external laser to the accelerating mode. But the open structure allowed for easier alignment and avoided the difficulties involved in transporting electrons through a narrow channel.

One consequence of using slower electrons was that for a similar laser wavelength, the accelerator's small-scale features—the grating teeth—would have needed to be one-third the size to match the speed of the accelerating wave to the particle velocity. But that challenge was surmountable. The MPQ team acquired custom-made single-sided gratings and used their laser to excite a harmonic mode with triple the frequency to accelerate the particles. Breuer observed a signal of accelerated electrons with 25 MeV/m accelerating gradient, which was comparable to that in conventional accelerators. But the new setup had the potential to significantly improve either by incorporating alternate structure choices or by operating at the fundamental harmonic, where the coupling between the laser field and the structure is stronger.

The SLAC/Stanford and FAU/MPQ demonstrations were both experimental firsts and demonstrated several interrelated concepts: that microstructure-based laser acceleration could be used for both subrelativistic and relativistic particles, that acceleration could be performed using harmonics of the laser wavelength, and that accelerating gradients meeting or exceeding those of conventional accelerators were possible in DLAs. It took about six months to analyze and write up the results, which appeared online in separate journals on the same day.[3,4]

Collaboration fosters innovation

The initial experiments garnered some attention in the press, and that coverage caused the phrase "accelerator on a chip" to gain traction. Building on the strong collaborations formed under LEAP and AXiS, in 2015 we assembled an international team of scientists and accelerator engineers with an array of expertise spanning accelerator science, photonics, laser technology, materials science, computer simulation, and nanofabrication. The team included six universities (Stanford, FAU, Purdue, UCLA, the Technical University of Darmstadt, and the University of Hamburg), three government labs (SLAC in the US; the German Electron Synchrotron, or DESY, in Hamburg, Germany; and the Paul Scherrer Institute in Switzerland), and one industrial partner (Tech-X). At the invitation of the Gordon and Betty Moore Foundation, the team submitted a proposal to bring laser acceleration from a demonstration of principle to a fully functional feedback-controlled particle acceleration device by the end of 2022. The proposal was reviewed and funded in 2015 for a five-year period.

We named the collaboration Accelerator on a Chip International Program, or ACHIP. The newly formed program was tasked with generating and controlling electron beams with a conceptually new technology and with much stricter tolerances than are required for more conventional approaches. We were proposing a radical change in the entire paradigm of future accelerators. The concrete goals of ACHIP are a 1 MeV particle accelerator on a chip and optical control of so-called transverse effects, which include extreme-UV (XUV) and x-ray photon generation from the wiggled electron beam and optical focusing. The vacuum chamber containing the electron source, electron

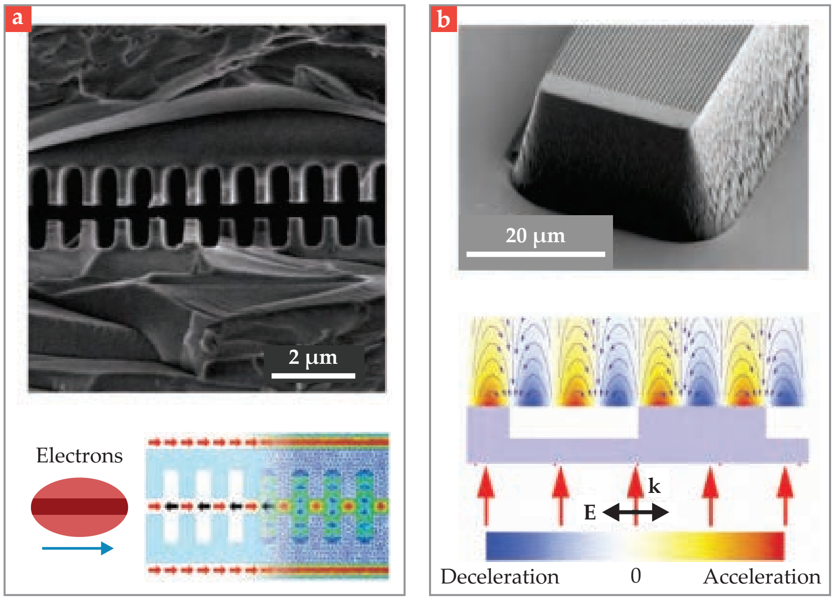


FIGURE 3. CHIP-SCALE ACCELERATORS made from fused silica were employed in two seminal laser-acceleration experiments in 2013. **(a)** The SLAC/Stanford University experiment used a dual-sided structure (top) with a 400-nm-wide channel for uniform field excitation (bottom). (Adapted from ref. 3.) **(b)** The experiment at Friedrich–Alexander University Erlangen–Nuremberg used an open-grating geometry (top) driven at the third harmonic, as illustrated by the contour plot (bottom). The grating's periodicity (purple) corresponds to three periodic repetitions of the longitudinal electric field. (Adapted from ref. 4.) In both experiments, periodic nanofabricated features match the velocity of the accelerating wave to that of the particles.

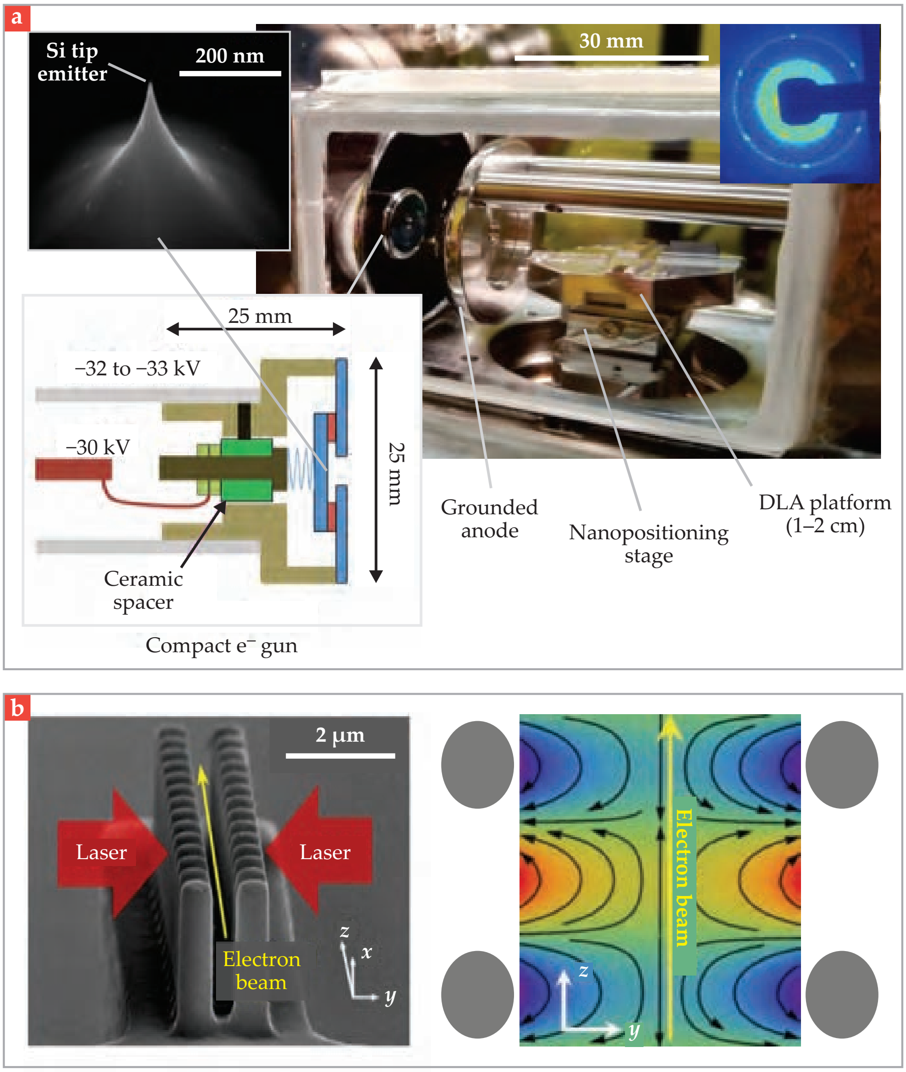


FIGURE 4. SILICON NANOFABRICATION allows for sophisticated structure designs that can be monolithically integrated on a chip. **(a)** A dual-pillar accelerator was incorporated into a 10-cm-long "glass box" test stand that is designed to achieve energy gains of 0.1–1 MeV. The device includes an integrated miniaturized electron gun (lower left) and silicon nanotip emitters (upper left). Such devices could enable ultracompact sources for electron diffraction studies, as illustrated by the diffraction pattern (upper right) obtained with a graphene target and a 57 keV beam. (Schematic adapted from ref. 5; silicon nanotip image courtesy of Andrew Ceballos; glass box and diffraction pattern images courtesy of Kenneth Leedle.) **(b)** The dual-pillar design consists of a colonnade of vertically standing silicon pillars. When driven in phase, two incident laser pulses excite a uniform accelerating mode between the pillars, shown here by simulated contour plots of the axial electric field E_z. (Adapted from ref. 18.)

optics, DLA chip, and detector should be able to fit in a shoebox. Parts of each goal have already been demonstrated, and the collaboration is on track to have a completed device in 2022.

The accelerator components fall into four major categories: the electron source, an injector and buncher that accelerate particles from subrelativistic to relativistic (MeV) energies, on-chip waveguides for controlled laser delivery, and relativistic speed-of-light acceleration. All the elements need to operate together in a final integrated device, which means reinventing them to be compatible with integrated Si photonics and lithographic fabrication. Five years ago that task appeared to be nearly impossible—a prerequisite for any project worth undertaking according to the motto of Edwin Land, inventor of the Polaroid camera.

The electron source must be compact and well suited for coupling to micron-sized devices. Progress there has largely exploited nanotip field-emission sources—tiny metallic or Si needles with tip radii on the order of tens of nanometers, from which electrons are emitted by photoemission excited by an ultrafast laser pulse. Nanotip emission sources have led to the development of centimeter-scale electron guns, which are now being integrated into compact tabletop demonstration accelerators, such as the "glass box" at Stanford,[5] shown in figure 4a.

An important aspect of the electron sources is that because of their tiny size, they have exceptionally high beam brightness—a measure of how small and well-collimated an electron beam can be made. The Si tips employed are quite robust and can last for days or weeks of continuous operation. As a tradeoff for their intense beam brightness, however, nanotip emission sources produce on the order of a few thousand electrons per incident laser pulse, compared with millions or billions of electrons for more conventional sources.

The accelerator naturally divides the electron bunches from the nanotip into smaller, optically spaced microbunches with durations that are less than 1 fs, a fraction of a laser cycle. The particles that successfully transition into a microbunch are "captured." The microbunch spacing is determined by the period of the accelerator—the spacing between grating teeth—and is therefore intrinsic to the design.

The particles naturally form microbunches as they get captured in the potential wells in the electromagnetic wave. The efficiency of the capture process can be enhanced by prebunching the particles. Recently, bunches with durations from 270 to 700 attoseconds have been measured and injected into a subsequent acceleration stage to perform fully on-chip bunching and net acceleration demonstrations.[6,7]

Successfully directing an electron beam through tiny accelerating devices is a major experimental challenge. For a sense of scale, the hollow acceleration channel inside the chip shown in figure 1 is just 400 nm wide and 1 mm long. Electrostatic lenses are incorporated into the gun to inject the electron beam into an accelerator channel. The lenses can focus the beam down to a submicron spot suitable for coupling into the narrow channel. The electrons are then transported over thousands of optical periods to reach substantial energy gains, from about 100 keV to 1 MeV.

Researchers have explored new innovative designs for accelerator structures, including replacing the grating-like structures used in the earlier experiments with free-standing rows of Si pillars separated by a narrow submicron channel, as shown in figure 4b. The dual-pillar design eliminates unnecessary material that could distort the incident laser pulse. It allows for highly symmetrical fields that can deflect, compress, or accelerate electrons in the channel by adjustment of the incident laser phases.[8]

The structures can focus and confine electrons inside the accelerating channel. They use the inherent laser-driven fields to enable a technique called alternating phase focusing, which was originally used in conventional RF proton accelerators and was reworked for use in photonic devices by researchers at the Technical University of Darmstadt.[9] Simulations of long-distance transport of the electron beam over multiple centimeters have been carried out using that scheme, and it serves as the basis for newly fabricated designs aimed at achieving MeV-scale acceleration in a single on-chip device with minimal particle loss.

Bringing a microaccelerator to light

Any accelerator design based on pulsed particle beams and pulsed electromagnetic energy must rely on a series of sequential acceleration stages. For a laser-driven accelerator, the length of each stage is determined by the interaction time of a particle bunch with a laser pulse. That time is set by their respective durations, typically on the order of 0.1–1 ps.

To maintain the phase-velocity match of the laser wave and the particle beam over many stages, the laser must be recoupled at each stage and its phase and amplitude carefully controlled. A compact and elegant solution is to couple the laser beam into on-chip photonic waveguides so the laser light can be split and precisely delivered to each successive segment of the accelerator. Photonic waveguides based on Si and silicon nitride materials are well developed and have been shown to support broad-bandwidth femtosecond-class laser pulses with up to 20 nJ per pulse, which are suitable for driving the individual stages of a chip-based accelerator.[10]

Researchers have carried out extensive optimization studies to evaluate the requirements for a complete multistage system based on laser delivery by waveguide.[11] The first integrated accelerator, which was recently demonstrated, coupled an external laser to an on-chip waveguide that fed into a custom-designed accelerator channel.[12] Inverse-design codes were used to optimize the photonic structure's design. The codes produced highly nonintuitive, almost organic-looking geometries (see figure 5) that were tailored to specific optimization parameters such as bandwidth and accelerating field.

Another consequence of accelerating subrelativistic electrons in an optical-scale device is that the particle energy changes along the length of the accelerator. The design velocity of the structure must therefore be varied along its length to accommodate the anticipated particle speed at each point along the accelerator. Two parallel methods can address that requirement. One is to vary, or chirp, the periodicity of the accelerator structure itself to match the local wave velocity to the electron speed.[13] The other is to imprint a custom phase profile onto the laser pulse.

By simultaneously chirping the laser pulse and tilting the wavefronts using dispersive optics, both phase and group velocities can be matched to achieve high-gradient acceleration

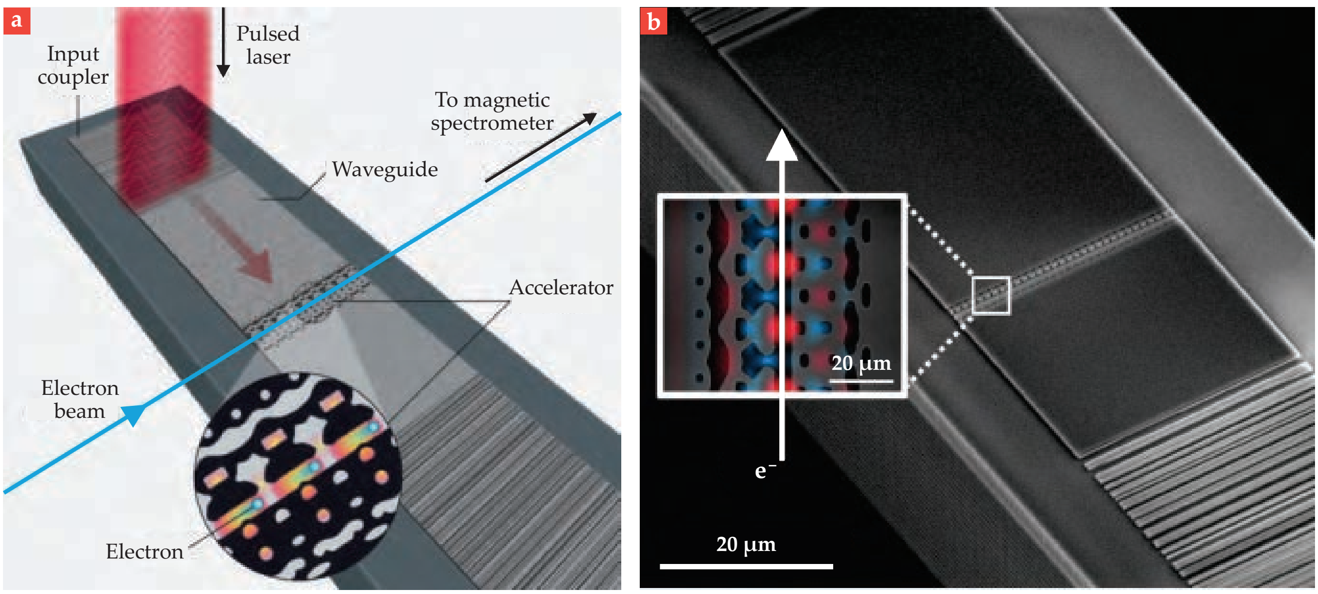


FIGURE 5. INVERSE DESIGN provides a powerful tool for producing highly optimized designs, such as this first demonstration of a single-stage integrated waveguide-coupled accelerator. **(a)** Shown here is a schematic layout of the experiment. **(b)** This electron micrograph image presents the as-built structure. The concept provides a key stepping-stone to future devices with multiple such stages driven by a network of phase-controlled waveguides. (Adapted from ref. 12.)

and laser-driven focusing over a centimeter-scale interaction length. That result illustrates a key aspect of the DLA approach, which is that beam dynamics can be controlled through careful manipulation of the laser phase.[14]

With tabletop integrated accelerator setups in operation, researchers are testing a new generation of structures that incorporate microbunching and laser-driven focusing. Over the next year, a series of planned experiments are expected to incrementally increase the total beam energy up to 1 MeV or higher. Once multi-MeV energies are reached, the particle velocities will become essentially constant at close to the speed of light.

A new set of relativistic-energy experiments are underway at facilities that have access to relativistic test beams. At the UCLA Pegasus Laboratory, moderately relativistic (4–8 MeV) electron beams are probing the limits of accelerating gradients and energy gains in DLA technology. Researchers there have demonstrated axial accelerating-field intensities up to 1.8 GV/m and an energy gain of 0.3 MeV.[14,15]

FIGURE 6. THE ACHIP COLLABORATION met at the German Electron Synchrotron (DESY) in Hamburg, Germany, on 19 September 2018. (Photo courtesy of Jann Wilken/DESY.)

New facilities are near completion at DESY in Germany and at the SwissFEL in Switzerland to test concepts relevant for beams with larger currents that will carry millions to billions of electrons per pulse and higher particle energies of 100 MeV and 3 GeV, respectively. Through a combination of innovative approaches that leverage the multi-institutional and interdisciplinary capabilities of the research teams, the promising results obtained over the past five years provide a toolbox of technologies to realize fully integrated on-chip accelerators that can be used for scientific, industrial, and medical applications.

Revolutionary devices

One of the most promising applications for DLA devices is compact electron sources for ultrafast science and electron diffraction studies. Trains of ultrafast and ultrabright electron bunches with subfemtosecond intrinsic bunch structure could allow resolution of electronic processes in both spatial and temporal domains, thereby enabling new tools for imaging light–matter interactions and for experiments in quantum electrodynamics.

Compact accelerators with target energies in the few-MeV range are attractive near-term candidates for use in medical dosimetry. Electron sources based on laser-driven on-chip accelerators could potentially fit on the end of an optical fiber, be placed on a scanning platform at the surface of a sample, or even get inserted into living tissue. A self-contained multi-MeV electron source based on integrated photonic particle accelerators could enable minimally invasive cancer treatments and adjustable-dose real-time deposition with improved dose control. One could envision an encapsulated micro-accelerator, built onto the end of an optical fiber, being placed at a tumor site using standard endoscopic methods. That could allow a doctor to deliver the same or higher radiation dose compared with what is provided by existing external-beam technologies, but with less damage to surrounding tissue.

The difference in bunch charge and duration for DLA electron beams compared with traditional RF accelerators points to the potential for future DLA-based light sources to generate attosecond pulses of XUV or x-ray radiation. Those pulses could probe matter on even shorter time scales than are possible today. Coherent attosecond radiation could also be produced using the same accelerator-on-a-chip operating principles. Proposed laser-driven dielectric undulators that wiggle the electron beam and produce radiation could be fabricated using photolithographic methods similar to those used to produce current DLA devices.[16]

Although a high-energy particle collider based on DLA technology is admittedly decades away, considerations for physics at the TeV scale indicate that DLAs have the potential to achieve the required luminosities with reasonable power consumption. The highest priority challenges for such future applications largely pertain to the transport of high average beam currents in the micron-scale apertures of DLA devices. But recent studies of extended structures with laser-driven focusing show promise for improved charge transport, capture efficiency, and preservation of beam quality.[9] They also support the prospect of parallel accelerating channels.[17]

Accelerating toward the future

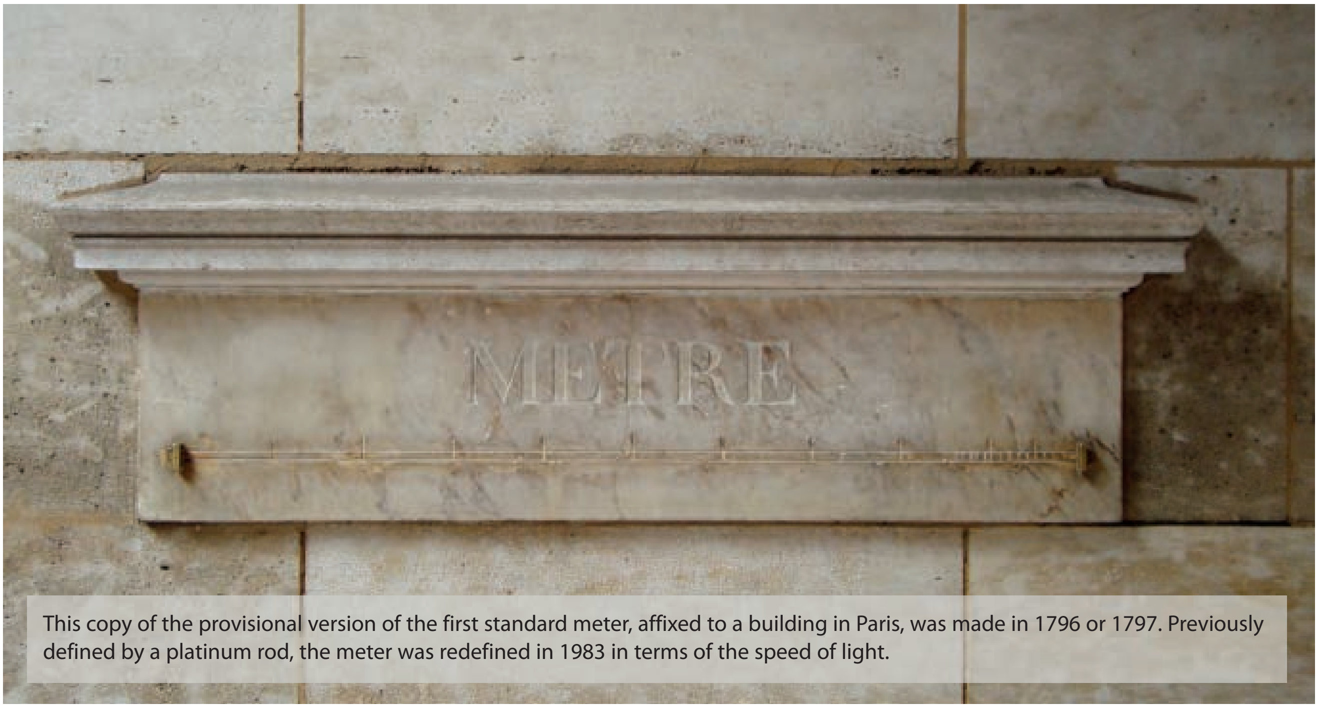
The idea of laser-driven accelerators was first proposed shortly after the demonstration of the laser. Now ACHIP is a worldwide collaboration (see figure 6) aimed at realizing an integrated laser-driven accelerator on a chip. The program melds the capabilities of universities and government laboratories and therefore has access to the facilities and talent needed to address key challenges in the endeavor. That collaboration enables projects to be handed off to national laboratories in a timely way so that they can handle the more costly engineering of a laser accelerator system. The emerging field of laser accelerators has engaged and trained scientists who can lead the development of new tools that will enable discoveries in areas of physics ranging from atomic and molecular to chemical, biological, and medical.

Fifty years ago the laser was a solution looking for a problem. Today it is a critical "stealth utility" that enables the technology of the modern world. The development of laser-driven attosecond electron sources may bring new capabilities in search of their own problems to solve. Fifty years from now, the DLA may be an essential tool in our lives—and perhaps even a stealth utility itself.

The authors thank current and past members of the ACHIP Advisory Board—Chan Joshi, Reinhard Brinkmann, Lia Merminga, and Tor Raubenheimer—for their guidance, and all members and students of the ACHIP collaboration for their many contributions. We thank Gary Greenburg for his early assistance and ongoing support. We also thank the Gordon and Betty Moore Foundation, NSF, the US Defense Advanced Research Projects Agency, the US Department of Energy, the German Federal Ministry of Education and Research, and the European Research Council for their current and past support.

REFERENCES

1. T. Plettner et al., *Phys. Rev. Lett.* **95**, 134801 (2005).
2. R. J. England et al., *Rev. Mod. Phys.* **86**, 1337 (2014).
3. E. A. Peralta et al., *Nature* **503**, 91 (2013).
4. J. Breuer, P. Hommelhoff, *Phys. Rev. Lett.* **111**, 134803 (2013).
5. T. Hirano et al., *Appl. Phys. Lett.* **116**, 161106 (2020).
6. N. Schönenberger et al., *Phys. Rev. Lett.* **123**, 264803 (2019).
7. D. S. Black et al., *Phys. Rev. Lett.* **123**, 264802 (2019).
8. K. J. Leedle et al., *Opt. Lett.* **40**, 4344 (2015).
9. U. Niedermayer et al., *Phys. Rev. Lett.* **121**, 214801 (2018).
10. S. Tan et al., *Opt. Lett.* **44**, 335 (2019).
11. T. W. Hughes et al., *Phys. Rev. Appl.* **9**, 054017 (2018).
12. N. V. Sapra et al., *Science* **367**, 79 (2020).
13. J. McNeur et al., *Optica* **5**, 687 (2018).
14. D. Cesar et al., *Commun. Phys.* **1**, 46 (2018).
15. D. Cesar et al., *Opt. Express* **26**, 29216 (2018).
16. T. Plettner, R. L. Byer, *Phys. Rev. ST Accel. Beams* **11**, 030704 (2008).
17. Z. Zhao et al., *Photonics Res.* **8**, 1586 (2020).
18. K. J. Leedle et al., *Opt. Lett.* **43**, 2181 (2018).



This copy of the provisional version of the first standard meter, affixed to a building in Paris, was made in 1796 or 1797. Previously defined by a platinum rod, the meter was redefined in 1983 in terms of the speed of light.

The history of a universal constant

The speed of light is one of the few universal constants in the universe, and the effort to measure it was one of the most enduring scientific problems in physics. In *Lightspeed: The Ghostly Aether and the Race to Measure the Speed of Light*, physicist John C. H. Spence travels through more than 2500 years of human history to reveal the lengths to which people have gone for an answer, from building a system of mirrors on the rooftops of Paris to spending eight years in India waiting to measure the transit of Venus. (Spoiler: It was cloudy that day.)

The subtitle bills the book as "the race to measure the speed of light," and although the term "race" may oversell the urgency of the enterprise, Spence does his best to keep things lively. Obtaining the measurement itself was fairly straightforward: 17th-century attempts using Jupiter's moons were surprisingly good, and 19th-century efforts with rotating mirrors got even closer to the value accepted today.

But that is just the starting point for what Spence calls an "adventure" that seeks to place the effort in the context of a bigger set of questions about the nature of light and its ability to move through space. Spence gleefully follows digressions—almost all of which are interesting—through topics as varied as stellar aberration, Foucault's pendulum, and Saturn's rings. He draws effectively on the work of prominent historians such as Olivier Darrigol and Bruce Hunt, albeit with an occasional misstep like referring to the French physicist Alfred Cornu by his first given name, Marie, which Cornu never went by.

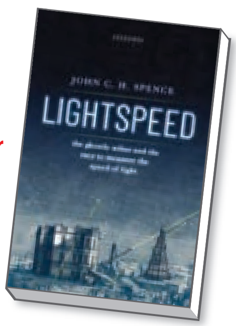
About halfway through the book, Spence shifts his focus from the experimental measurement of light speed to a different, more theoretical set of questions about the nature of how light moves. That portion centers on Albert Michelson, who began his career with efforts to measure the speed of light. But Spence admits that those early experiments by Michelson are not terribly interesting and mostly refined the techniques used by Hippolyte Fizeau and Léon Foucault, so he moves instead to Michelson's next project: measuring the "ether wind" caused by Earth's motion.

It's a well-worn myth that the Michelson–Morley experiment, which was unable to detect any motion of Earth relative to the ether, inspired Albert Einstein to reject the existence of the ether entirely and propose instead the special theory of relativity. Spence does not quite fall into that trap, as he acknowledges that Einstein saw his own work as addressing the theoretical problem of reconciling Maxwell's equations of electrodynamics with the physics of motion.

But he does place the Michelson–Morley experiment at the center of the action and frames his narrative around it. Spence quotes a 1931 speech by Einstein, in which he credited Michelson for having "paved the way for the theory of relativity," although one may imagine

Lightspeed
The Ghostly Aether and the Race to Measure the Speed of Light

John C. H. Spence
Oxford U. Press, 2020. $32.95



Einstein was being polite to Michelson, as Michelson had invited him to give that speech. Spence also claims that Einstein referred to Michelson's experiment in his 1905 paper "On the electrodynamics of moving bodies," but that is not the case. Although Einstein did allude to "unsuccessful attempts to discover any motion of the earth relatively to the 'light medium,'" historians like Gerald Holton have shown that he was not specifically referring to Michelson.

Spence points out that we now live in a moment where the concept of measuring the speed of light has been turned on its head. Einstein's work elevated *c* to a fundamental constant of nature, and in 1983 the General Conference on Weights and Measures gave it an exact value. Previously, the speed of anything had been measured by the distance—traditionally defined by the length of a platinum bar preserved in a vault in Paris—that it traveled over time, traditionally defined as a fraction of the day. But the conference has scrapped the platinum bar and effectively defined the speed of light as exactly 299 792 458 meters per second, meaning that the speed of light now determines the length of a meter.

Spence ends the book by asking whether there is any possibility of information moving faster than the speed of light. He then presents a lengthy discussion of Bell's theorem and spooky action at a distance, but notes that the question remains unresolved. He also mentions that the process of writing the book changed his own position on the matter.

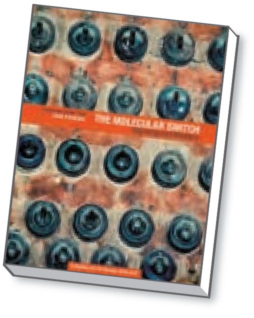
Lightspeed has the tone of an exuberant physics lecture, complete with a lab exercise at the end that involves measuring the speed of light with pizza dough. The book's strength is its clear and thorough explanation of the underlying physics, which does not necessarily assume any prior knowledge of the subject but quickly accelerates to a high level. For that reason, it is hard to imagine someone sticking with Spence if they haven't already decided that they like physics. But for those of us who do, *Lightspeed* is an animated account that vividly evokes the numerous and often outsized personalities who contributed to figuring out just how fast light travels.

Theresa Levitt
University of Mississippi
Oxford

Many proteins can be analogized to simple switches with on and off states, like the ones present on the US Capitol's switchboard, pictured here in 1959.

LIBRARY OF CONGRESS, LC-U9-1924-25

Biology unified by physics



The Molecular Switch
Signaling and Allostery
Rob Phillips
Princeton U. Press, 2020. $85.00

Biology is often taught with diagrams. But you can also teach it with equations, the language of physics. In his new book, *The Molecular Switch: Signaling and Allostery*, biophysicist Rob Phillips does just that, explaining allostery—a biological enigma—using the formalism of statistical physics.

Phillips is a professor of biophysics, biology, and physics at Caltech. He has authored or coauthored several books that creatively combine concepts from biology, physics, and materials science, including *Crystals, Defects and Microstructures: Modeling Across Scales* (2001); *Physical Biology of the Cell* (2nd edition, 2012); and *Cell Biology by the Numbers* (2016). *The Molecular Switch* is another fantastic book in that vein. It is an elegant demonstration of how physics can be used to explain a biological concept and unify a wide range of biological phenomena. The book is a great introduction to biophysics for both physicists and biologists and a versatile reference for advanced undergraduate and graduate biophysics courses.

Proteins are the building blocks and functional molecular units of living cells. For cells to operate properly, proteins must function in a controlled manner. But how do they do that? Richard Feynman acknowledged in his 1963 lectures on physics that "everything that living things do can be understood in terms of the jigglings and wigglings of atoms." His maxim also holds true for proteins. They fold into elaborate three-dimensional structures and can interact with other molecules. Those intra- and intermolecular changes are dynamic and affect a protein's function over time.

If its activity changes on the binding of a small molecule, a protein is said to be allosteric, a term originally coined by Jacques Monod and François Jacob in 1961. Many, if not all, proteins are expected to be allosteric. Because its consequences are so profound, allostery has been called the second secret of life after the genetic code.

Despite its broad impact in the field, allostery is explained rather briefly in most biology textbooks, with just one example—namely, the cooperative binding of oxygen in hemoglobin, the protein that carries the molecule in our blood. The primary achievement of Phillips's new book is twofold: It expands the discussion of allostery to almost everything in nature and suggests a new paradigm for studying it.

As illustrated on the book's cover, that new paradigm is based on switches. Because many proteins can be likened to switches with on and off or active and inactive states, cellular functions can be viewed as input–output responses of those molecular switches. The switch analogy allows biophysicists to ignore the microscopic details of different allosteric molecules and analyze protein functions in terms of discrete states through the formalisms of equilibrium statistical physics.

Using that paradigm, Phillips aims to communicate with audiences from both biology and physics. He tactfully achieves that goal through his straightforward and logical organization. In chapters 1 and 2, Phillips explains allostery from the perspectives of biology and physics so that all readers are on the same page. He lays out basic statistical-physics concepts so that his model can be easily derived. In chapter 3, he uses the simple example of ion channels to demonstrate how statistical physics can be used to describe a molecular switch and experimentally measurable behaviors such as leakiness and dynamic range.

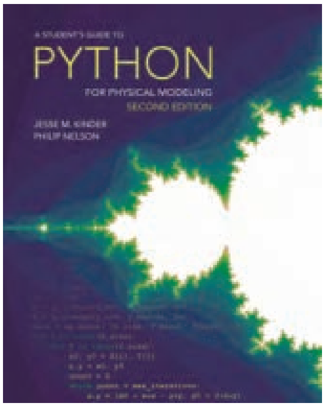
Phillips then applies that formalism with depth, breadth, and clarity to various biological topics, including the swimming of bacteria in response to a chemoattractant, cell signaling in such processes as photon detection in the eye, and gene regulation by transcription factors or nucleosomes. Those examples are presented not as information to be memorized but as a series of applications of the mathematical framework presented in the earlier chapters. Phillips also takes readers over multiple scales of biology, from animals to cells and molecules, and thus highlights the significance and universality of allostery in nature.

Teaching a biophysics course can be challenging because students have mixed backgrounds and expectations. Students from physics may be overwhelmed by biological details, which can prevent them from seeing biophysics' bigger picture. On the other hand, students from biology might feel frustrated by physics equations. *The Molecular Switch* serves students from both fields, who will be happy to see that a simple physical model can explain so many different and seemingly disparate phenomena. Another educational benefit of the book is that it teaches students how to develop effective theories to describe biological phenomena, which will be crucial for future collaborations between physicists and biologists.

The concept of allostery has been around for several decades. But who would have thought it could be used to create a physical law that unifies biology? Phillips does just that in *The Molecular Switch*. Anyone who likes the beauty of unification will appreciate his ingenious approach.

Sangjin Kim
University of Illinois at Urbana-Champaign

NEW BOOKS & MEDIA



A Student's Guide to Python for Physical Modeling

Jesse M. Kinder and Philip Nelson

Princeton U. Press, 2021 (2nd ed.). $75.00

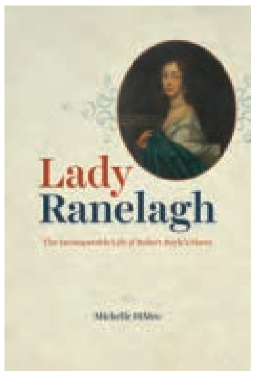
Long gone are the days when experimentalists' sole tools were laboratory setups, telescopes, and particle accelerators. Today, computer modeling is crucial to experimentation in fields from biophysics to astronomy. In *A Student's Guide to Python for Physical Modeling*, two physicists who use computing heavily in their daily workflow, Jesse Kinder and Philip Nelson, present students with the basics of Python programming for physics applications. The book takes a conversational tone and assumes no prior knowledge of programming, meaning that it can be used as a textbook for a physics programming course or for self-study. Students go from learning how to import libraries of functions into their Python environments to coding complex programs that can execute random walk simulations and create data visualizations. The book even includes appendices that walk students through the process of installing Python on various operating systems. –RD

Lady Ranelagh

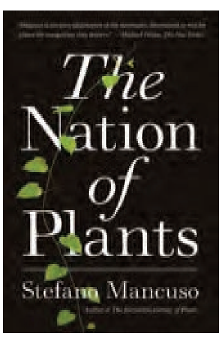
The Incomparable Life of Robert Boyle's Sister

Michelle DiMeo

U. Chicago Press, 2021. $45.00



One of the most respected and influential women of her time, Irish-born Lady Ranelagh was a 17th-century natural philosopher, intellectual, and member of London's Hartlib circle, a precursor to the Royal Society. Ranelagh's works were never published nor were her manuscripts preserved, so she is much less well known than her brother, the distinguished scientist Robert Boyle. To address this, historian Michelle DiMeo has written the first full-length biography of Ranelagh by gleaning details of her life from her correspondence and the archives and writings of her relatives and contemporaries. The result is a detailed account of this notable woman, her work, and her close, collaborative relationship with her brother Robert—set against the backdrop of the turbulent politics of the times, including the Irish and English civil wars. –CC

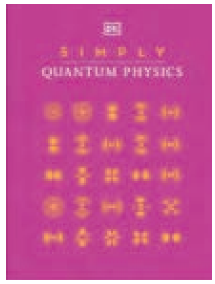


The Nation of Plants

Stefano Mancuso, trans. by Gregory Conti

Other Press, 2021. $21.99

In this imaginative call to action, Italian botanist Stefano Mancuso addresses current environmental crises like climate change and decreasing biodiversity from the perspective of the "oldest and most populous" nation on the planet—plants. Acting as their representative, Mancuso presents a constitution that spells out eight fundamental pillars on which rest the lives not only of plants but of all living beings. That constitution launches a discussion that juxtaposes plants' importance as a carbon dioxide sink and a source of food, energy, building materials, and other resources against humans' wastefulness and anthropocentrism. –CC



Simply Quantum Physics

DK, 2021. $16.99

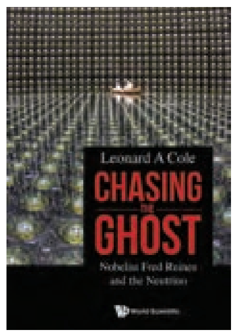
An illustrated reference book aimed at the general reader, *Simply Quantum Physics* is an introductory overview of the quantum world. Diagrams and schematics appear on almost every page, and the accompanying text is brief and nontechnical. In addition to discussing the structure of the atom and of subatomic particles and the fundamental forces holding them together, the book covers some quirky quantum-level properties of matter, such as wave–particle duality. It also describes electron microscopes, LEDs, and unusual phenomena like quantum teleportation. Particle physicist Ben Still served as consultant editor; he is the author of several popular science books, including an illustrated reference book based on LEGO building blocks, *Particle Physics Brick by Brick* (see PHYSICS TODAY, May 2018, page 62). –CC

Chasing the Ghost

Nobelist Fred Reines and the Neutrino

Leonard A. Cole

World Scientific, 2021. $68.00



Physicist Frederick Reines, who won the 1995 Nobel Prize in Physics for his work with Clyde Cowan to experimentally detect the neutrino, is the subject of *Chasing the Ghost*, the first book-length study of his life. Authored by Reines's cousin Leonard A. Cole, the book is a unique mix of memoir and biography. Although Cole's connection to Reines allows him to add context and flavor to sections detailing Reines's childhood, it also leads readers into Reines's teenage diaries, digressions more appropriate for a family history. Nevertheless, the book will be valuable to scholars interested in the life of a 20th-century experimental pioneer. –RD PT

NEW PRODUCTS

Focus on photonics, spectroscopy, and spectrometry

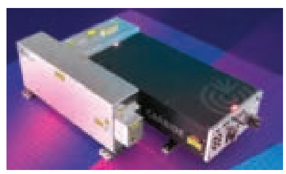
The descriptions of the new products listed in this section are based on information supplied to us by the manufacturers. PHYSICS TODAY can assume no responsibility for their accuracy. For more information about a particular product, visit the website at the end of its description. Please send all new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

Andreas Mandelis

Tunable laser for Raman spectroscopy

Hübner Photonics has announced that its C-Wave laser series has qualified for use with the TriVista spectrometer system from S&I Spectroscopy & Imaging. The companies expect that the combination of Hübner's widely tunable CW single-frequency lasers, which cover wavelengths from 450 nm to 3.5 μm, with S&I's triple-grating TriVista will lead to advances in resonance Raman spectroscopy and microscopy, especially for the challenging low-frequency Raman range of less than 10 cm^{-1}. Most tunable lasers require elaborate filtering to suppress unwanted amplified spontaneous emission (ASE) that can cover the weak Raman signals. During qualification, the C-Wave revealed no detectable ASE. The C-Wave's narrow linewidth, stable emission frequency, and high spectral purity allow for recording low-frequency Raman bands at or below 5 cm^{-1} from the excitation wavelength even without filtering the excitation beam. ***Hübner Photonics Inc***, *2635 N 1st St, Ste 202, San Jose, CA 95134, https://hubner-photonics.com*

Laser source for nonlinear microscopy



Light Conversion developed its Cronus-3P laser source specifically for advanced nonlinear microscopy. Based on optical parametric amplifier technology, the Cronus-3P provides μJ-level sub-85 fs pulses at repetition rates of up to 2 MHz. It is tunable from 1.25 μm to 1.8 μm, which covers the biological transparency windows at 1.3 μm and 1.7 μm for three-photon microscopy. The source features integrated group-delay-dispersion control to ensure optimal pulse duration at the sample, automated beam resizing and collimation, and optional automated beam steering for laser-pointing stability. ***Light Conversion***, *Keramiku St 2B, LT-10233 Vilnius, Lithuania, https://lightcon.com*

Chromatographic analysis of metal-sensitive compounds

Waters has introduced its Arc Premier System, the first liquid chromatography system optimized for chromatographic separations on 2.5–3.5 μm columns to also feature Waters's new MaxPeak High Performance Surfaces (HPS) technology. The hybrid organic–inorganic surface technology, exclusive to the company's MaxPeak Premier Systems and Columns, forms a barrier between the sample and metal surfaces. Without compromising performance, it mitigates or eliminates nonspecific adsorption and maximizes reproducibility and efficiency of separations. The new system complements the company's MaxPeak Premier Columns to eliminate time-consuming and costly passivation and deliver high-quality, accurate sample data. The removal of analyte-to-metal interactions can result in up to a fivefold improvement in detector sensitivity, depending on the degree of metal sensitivity, and a tenfold improvement in assay-to-assay precision. ***Waters Corporation***, *34 Maple St, Milford, MA 01757, www.waters.com*

Large-array cameras for astronomy

Teledyne Princeton Instruments, a business unit of Teledyne Digital Imaging, has brought to market its Cosmos large-format, backside-illuminated CMOS cameras optimized for astronomy. Specific applications include time-domain astrophysics, orbital-object tracking, and exoplanet research. The cameras incorporate Teledyne's proprietary LACera image-sensor technology, which delivers greater than 90% quantum efficiency for high sensitivity and proprietary low-noise architecture with up to 18-bit readout. According to the company, that level of performance was not previously available in wafer-scale sensors. Cosmos is offered in three sensor sizes: 3000 × 3000 pixels, 6000 × 6000 pixels, and 8000 × 8000 pixels. It provides more than 50 frames/s for capturing dynamic events, global shutter, glow-reduction technology, 0.7 e^- read noise for detection of faint objects, and deep cooling to ensure low dark current. ***Teledyne Princeton Instruments***, *3660 Quakerbridge Rd, Trenton, NJ 08619, www.princetoninstruments.com*

High-power visible Raman fiber amplifiers

MPB Communications now offers single-frequency (SF) visible Raman fiber amplifiers (VRFAs) that achieve output powers of up to 4 W in the 525–650 nm spectral range. The VRFAs provide a diffraction-limited beam with excellent pointing stability, low-phase and low-intensity noise, and minimal spectral broadening. The system consists of a polarization-maintaining Raman fiber amplifier (RFA) pumped by an ytterbium fiber laser. They are housed in a compact, air-cooled platform. High-power SF emission from the RFA is frequency-doubled by a single-pass, second-harmonic generator (SHG). To eliminate vibration transfer and simplify experimental setup, the SHG and RFA can be mounted away from the pump laser. The SHG output can be either fiber coupled or free space. The higher-power VRFA systems address the evolving needs of quantum physicists studying quantum degeneracy, atomic clocks, ultracold atomic clouds, and trapped Rydberg ions. They also have applications in spectroscopy, atomic interferometry, optics, lithography, and gravitational measurements. ***MPB Communications Inc****, 147 Hymus Blvd, Pointe-Claire, QC H9R 1E9, Canada, www.mpbc.ca*

Raman system with remote probes

The latest version of Renishaw's Virsa Raman system includes features designed to expand the use of Raman spectroscopy to new samples, applications, and environments. Remote fiber-optic probes let users analyze samples away from the laboratory microscope. LiveTrack focus-tracking technology and the new Monitor software module allow for easy real-time analysis on large samples that have irregular surfaces, change shape as they undergo phase changes, or move. Windows-based Raman Environment (WiRE) software, version 5.5, adds two features that, used with the Monitor module, complement the Virsa system: Live reaction monitoring allows users to process and analyze a constant flow of Raman data in order to observe changing chemical concentrations or other sample properties; the partial least squares (PLS) analysis module allows users to generate and test PLS models and predict values in real time for materials that exhibit a spectral change, such as in concentration or crystallinity. ***Renishaw plc****, 1001 Wesemann Dr, West Dundee, IL 60118, www.renishaw.com*

Updated electronics for spectrometers

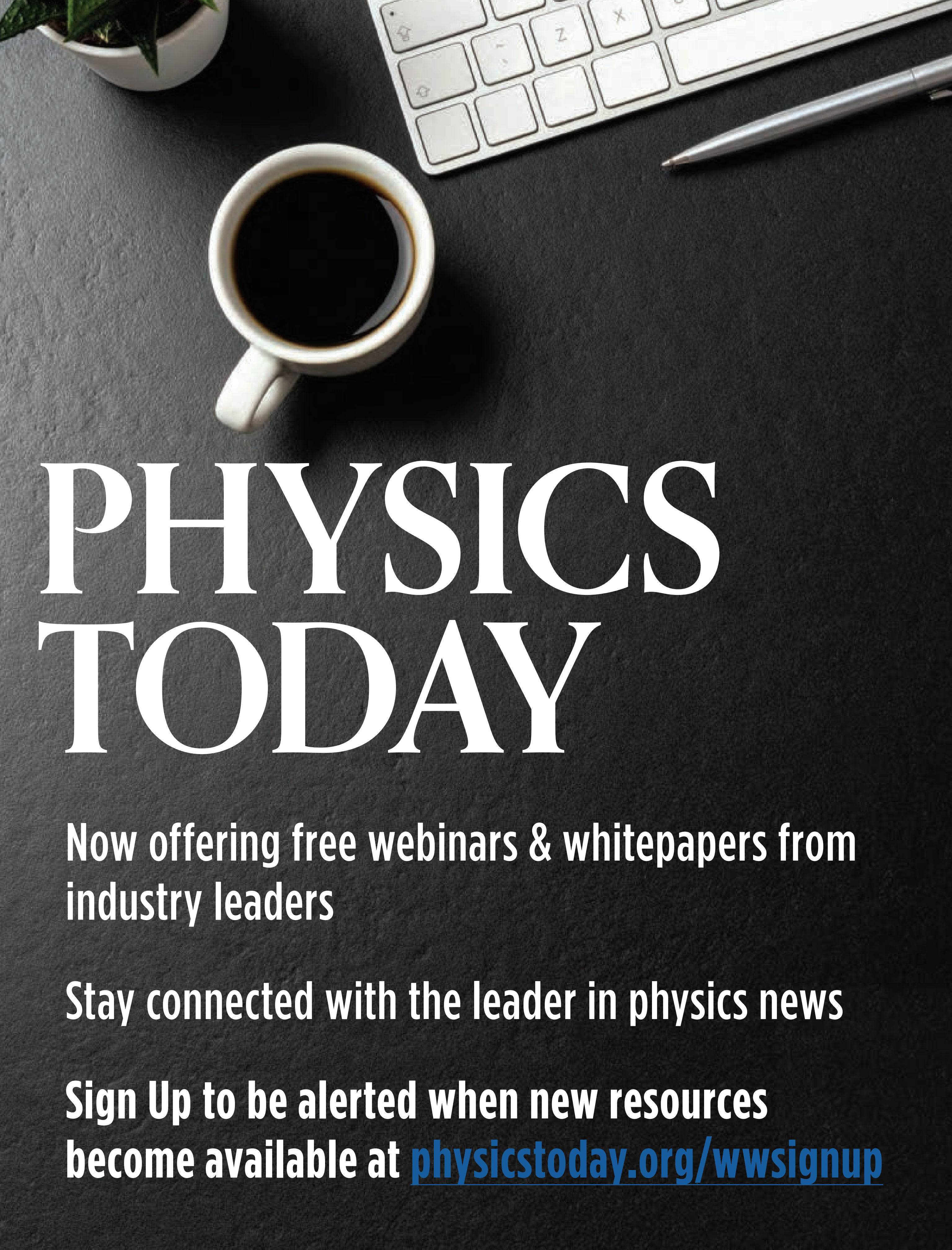
Ibsen Photonics has presented new electronics for its Freedom 305 and 315 OEM spectrometers. The electronics convert the analog video signal from the diode-array detector to a robust digital signal that can be read out over a standard interface such as USB or serial peripheral interface (SPI). The digital image sensor board DISB-315 offers a frame rate that is six times as fast as the previous versions. It supports the low-noise S10420 BT-CCD detector arrays from Hamamatsu, features programmable lamp and shutter control, and comes with the same SPI as the company's other DISBs, which makes it easy to integrate into instruments. If the spectrometer requires the use of a standard FTDI FT4222H chip, a compact DISB–USB bridge can be added and stacked on top of the DISB-315. ***Ibsen Photonics A/S****, Ryttermarken 17, DK-3520 Farum, Denmark, https://ibsen.com*

LED illuminator

Excelitas has added the X-Cite NOVEM to its fluorescence illumination product line. The LED illuminator is suitable for complicated imaging applications that require high excitation power and individual wavelength control. Those applications include slide scanning, live-cell imaging, fluorescence *in situ* hybridization, ratiometric imaging, and general fluorescence microscopy. The light-guide-coupled, nine-channel, wavelength-switching X-Cite NOVEM illuminator offers spectral ranges for applications from 340 nm to 785 nm. Preinstalled filters simplify system setup, and the device's high power reduces scan time for multiplex imaging. Available in four standard wavelength combinations, the X-Cite NOVEM features individual LED control, efficient cooling, and quiet operation, even when running at full capacity. Excelitas's patented LaserLED Hybrid Drive technology provides increased excitation in the 500–600 nm band range. ***Excelitas Technologies Corp****, 200 West St, 4th Flr E, Waltham, MA 02451, www.excelitas.com*

Hyperspectral light-field cameras

With the release of its Ultris X20 and X20 Plus models, Cubert has upgraded its 3D hyperspectral Ultris cameras to cover a wavelength range of 350–1000 nm. The innovative cameras, which are based on light-field technology, now offer UV, visible, and near-IR coverage. That capability may enable such applications as plant-water detection for vegetation analysis, real-color characterization, bathymetry, and water-quality and fluorescence analysis. Weighing less than 350 g, the lightweight Ultris X20 is also suitable for mapping applications by unmanned aerial vehicles. It offers scan rates of more than 3200 lines/s, a constant full width at half maximum of 10 nm across all channels, more than 160 spectral bands, and a native image resolution of 410 × 410 pixels. The high-resolution Ultris X20 Plus features a pan sensor, which allows the camera to achieve an image resolution of more than 1800 × 1800 pixels. ***Cubert GmbH****, Science Park II, Lise-Meitner Str 8/1, D-89081 Ulm, Germany, https://cubert-gmbh.com* PT


PHYSICS TODAY
Now offering free webinars & whitepapers from industry leaders
Stay connected with the leader in physics news
Sign Up to be alerted when new resources become available at physicstoday.org/wwsignup

Zdeněk Herman

In recent years physical chemist Zdeněk Herman, who died on 25 February 2021 in Prague, had been increasingly preoccupied with the startling opening of Charles Eliot's book *John Gilley of Baker's Island*: "To be absolutely forgotten in a few years is the common fate of mankind. . . . With the rarest exceptions, the death of each human individual is followed in a short time by complete oblivion, so far as living human memories are concerned." Its message was among Zdeněk's reasons for writing his so-far-unpublished memoir, which is about what, in his view, is memorable about others—and the network of relationships that bonded a community that grew out of a shared interest in chemical-reaction dynamics.

That community, of which Zdeněk was a cherished founding member, has been recorded not only in his prose but also in his countless drawings, made with the alacrity and accuracy of a master. His achievements as an artist—draftsman, painter, sculptor—compare to his scientific ones; indeed, the two are connected by his extraordinary dexterity, key to crafting and running fine scientific apparatus.

Born on 24 March 1934 in Libušín, Czechoslovakia, Zdeněk graduated from Prague's Charles University in 1957 with a diploma in radiation chemistry. His adviser, František Běhounek, had a second career as an author of adventure and science fiction books. Zdeněk, a dedicated Boy Scout, had read them voraciously, and meeting his idol was a dream come true.

He then went to the Institute of Physical Chemistry of the Czechoslovak Academy of Sciences (a predecessor of the J. Heyrovský Institute of Physical Chemistry and Electrochemistry). There, Vladimír Čermák introduced Zdeněk to mass spectrometry and the physics and chemistry of ions. Zdeněk's 1963 PhD thesis, done under Čermák, dealt with reactions of electronically excited ions in the ion source of a homemade mass spectrometer. The technique of studying ion–molecule reactions was inspired by Zdeněk's serendipitous discovery of a facile hydronium ion formation in the ion source after he did something he was *not* supposed to do—increase the source pressure (as it put the filament in danger of burning).

Ion–molecule reactions would become the dominant theme of Zdeněk's career. With a homemade multichannel jet as a molecular-beam source in his pocket, Zdeněk joined Richard Wolfgang at Yale University in 1964 to study the dynamics of ion–molecule reactions in crossed molecular beams under single-collision conditions. With several collaborators, Zdeněk and Wolfgang established the basic mechanisms of ion–molecule collisions and their dependence on the energy disposal. Zdeněk was in the US with his family in August 1968 when the Soviets invaded Czechoslovakia to crush the Prague Spring. Despite several offers his colleagues generated to keep Zdeněk in the US, the Hermans returned to Prague in 1969.

Throughout the 1970s, travel out of Czechoslovakia was severely restricted, and the small laboratory that Zdeněk shared with Čermák became a mecca for Western scientists traveling to Eastern Europe. More than 250 signatures of the world's best-known atomic, molecular, and chemical physicists decorated the ceiling of the Čermák–Herman laboratory.

Following the Velvet Revolution in 1989, Zdeněk took part in reconstructing the scientific life of his country and quickly became overloaded with administrative responsibilities, among them codirecting the Heyrovský Institute and cofounding the country's first research-grant agency. He also received a professorship at the Institute of Chemical Technology in Prague in 1996.

All that left little time for the laboratory, which, ironically, filled up with students anyway. Much of Zdeněk's research in the 1990s concerned the dynamics of single-charge transfer from doubly charged ions. Later, his research interests moved to the dynamics of ion–surface interactions.

Over the years, Zdeněk became the heart and soul of the biennial European Conference on the Dynamics of Molecular Systems (MOLEC) meetings, for which he designed and made a beautiful chairman's scepter. His key role was recognized in 2014 with the establishment of the Zdeněk Herman MOLEC Young Scientist Prize, for which he later designed its gold medals.

On his 80th birthday, his friends and colleagues established the Zdeněk Herman Jubilee Fund, which received more than 50 contributions. He used the fund to establish a prize for freshly minted Czech PhDs working in chemical physics and mass spectrometry.

HERMAN FAMILY/PRAGUE

Zdeněk Herman

For half a century, Zdeněk created drawings, paintings, and sculptures of friends and colleagues. In 2017 he published a 276-page book of his legendary art, noting, "In a way, it is a portrait of the generations of people that formed our years in science. . . . Perhaps this collection will keep the memory of not only their names, but also of their faces." The three of us all appear in the book. One beautiful work absent from it is a 50-cm-high wooden sculpture of the Madonna, which Zdeněk gave one of us (Herschbach) and which brings back wonderful memories of Zdeněk.

The international physical chemistry and chemical physics community has lost a remarkably gifted scientist, brilliant artist, and revered friend, who has done much to make our science a convivial and congenial adventure.

Dudley Herschbach
Harvard University
Cambridge, Massachusetts
Peter Toennies
Max Planck Institute for Dynamics and Self-Organization
Göttingen, Germany
Bretislav Friedrich
Fritz Haber Institute of the Max Planck Society
Berlin, Germany

Thomas James McIlrath

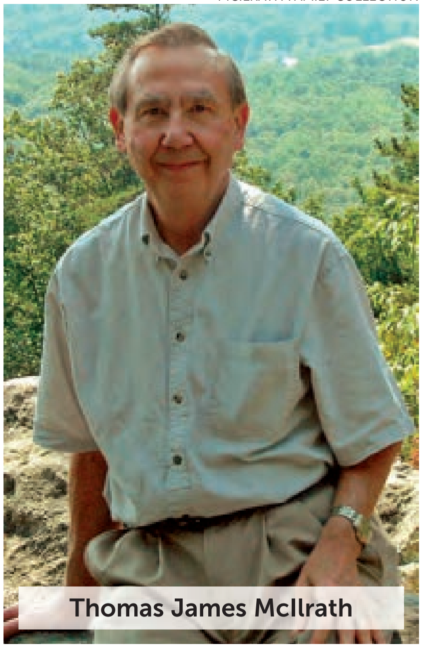
Thomas James McIlrath, former treasurer and publisher of the American Physical Society (APS), passed away on 20 December 2020 of complications from Parkinson's disease. In addition to being a skilled fiscal manager who maintained APS's financial health in a period of rapid economic change, Tom was an outstanding atomic, molecular, and optical (AMO) physicist. He spent countless hours volunteering with the Optical Society. He also was a faithful colleague and friend and a steadfast steward to his family and neighbors.

Tom was born in Dowagiac, Michigan, on 10 May 1938 and received his BS degree in physics at Michigan State University in 1960. His graduate work at Princeton University, done under the guidance of Thomas Carver, was on optical pumping of atomic hydrogen. After earning his PhD in 1966, he spent a year at Oxford University on a NATO postdoctoral fellowship. For the next six years, he pursued atomic spectroscopy research at the Harvard College Observatory while lecturing in the university's astronomy department.

Tom joined the faculty of the Institute for Fluid Dynamics and Applied Mathematics (an antecedent to today's Institute for Physical Science and Technology) at the University of Maryland (UMD) in 1973. During his time there, he was also a part-time staff researcher at the National Bureau of Standards (now NIST), where he collaborated extensively with Thomas Lucatorto and others. As part of Tom's fruitful UMD–NIST association, he, Lucatorto, and James Roberts developed a vacuum-ultraviolet (VUV) window that could sustain a pressure differential of seven orders of magnitude while transmitting 50% of the light. That work earned the three an *Industrial Research & Development* I-R 100 award (the precursor to the R&D 100 Awards presented by *R&D World*) in 1980.

With Lucatorto, McIlrath discovered superelastic collision-assisted resonant ionization of atomic vapors that enabled VUV absorption measurements of long columns of atomic ions (cold plasmas). McIlrath and Lucatorto received a Department of Commerce Silver Medal Award for their research in plasma studies and spectroscopy. One of us (Hill) worked closely with McIlrath, first as a postdoc at NIST with some of the ion columns and later as a colleague at UMD. McIlrath also provided valuable input in the 1980s and 1990s that helped lay the groundwork for UMD, NIST, and the Laboratory for Physical Sciences to create the Joint Quantum Institute.

MCILRATH FAMILY COLLECTION



Thomas James McIlrath

In 1986 Tom spent a sabbatical at Bell Labs, where he worked closely with another of us (Freeman), Philip Bucksbaum, and Howard Milchberg, then a postdoc. That was shortly after the discovery of above-threshold ionization by intense laser pulses, which at the time was not well understood. In the late 1980s, McIlrath, Freeman, and Bucksbaum wrote a series of papers describing the ionization of atoms and molecules in intense laser fields, an important step in the development of high-harmonic and attosecond generation in the 1990s and into the new millennium.

Between 1993 and 1995, Tom served at NSF as a rotator—a program director on temporary leave from one's home institution—in which he oversaw the AMO experimental program. In addition, he and Lawrence Goldberg (in the Engineering Directorate) cochaired a major 1994 NSF workshop, "Optical Science and Engineering: New Directions and Opportunities in Research and Education." It led to an NSF-wide initiative on multidisciplinary research in optical science and engineering that ultimately expanded research and education opportunities in those fields. Upon returning to UMD, Tom served as associate dean for research and graduate studies before retiring and becoming the treasurer and publisher at APS in 1996.

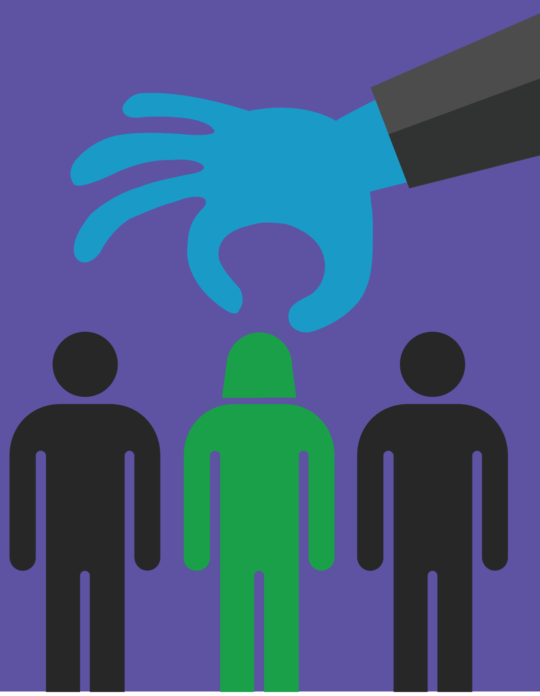
His achievements at APS included helping to steer the society into electronic publishing and to finance APS journals in a fairer and more equitable manner. For example, Tom started the tiered structure in which larger, research-intensive institutions began bearing more of the cost burden than non-PhD-granting, smaller institutions. Another legacy of his 10-year leadership—he stepped down in 2006—was his implementation of cost-saving measures, including online submissions, that led to dropping journal prices in the mid 2000s even as the size of the journals grew. Through it all, Tom was able to maintain the fiscal health of APS in a volatile economic period.

Tom faced a profound personal challenge when his wife, Valerie U. Hoy, was diagnosed with an incurable medical condition while he was in graduate school. His lifelong dedication to her and their two daughters, Christine and Laura, was exemplary, as was his service as deacon and elder of Riverdale Presbyterian Church. After his diagnosis with Parkinson's, he and Valerie moved to Asbury Methodist Village (AMV), a continuing-care retirement community in Gaithersburg, Maryland. In his typical altruistic fashion, he became very active, contributing to the AMV newsletter and interacting with AMV's department of pastoral care until his death.

We gratefully acknowledge important contributions from Christine McIlrath Lehnigk, Tom Lucatorto, Denise Caldwell, Larry Goldberg, Barry Schneider, Kate Kirby, Mike Stephens, David Voss, Phil Bucksbaum, and Howard Milchberg.

Wendell T. Hill III
Jan V. Sengers
University of Maryland
College Park

Richard R. Freeman
University of Washington
Seattle PT



QUICK STUDY

Eve Vavagiakis is a postdoctoral researcher in physics at Cornell University in Ithaca, New York. **Thomas Bachlechner** is head of artificial intelligence at MeetElise and **Matthew Kleban** is a professor of physics at New York University, both in New York City.



Is the electric potential physical?

Eve M. Vavagiakis, Thomas C. Bachlechner, and Matthew Kleban

In the yet-unmeasured electric Aharonov–Bohm effect, an electric potential influences the quantum mechanical wavefunction of charged matter—even in regions where the electric field vanishes.

More than a century after James Clerk Maxwell first elucidated the phenomenon of electromagnetism, the implications of his famous formulas are still not fully understood. Although Maxwell's equations contain only electric and magnetic fields, they can be conveniently expressed in terms of electric and magnetic potentials, quantities whose spatial and temporal variations determine the field strengths. In theory, the potentials themselves have no significance. All the physics is contained in the forces exerted on charged particles, and because those forces are directly proportional to the field strengths, they vanish where the field strengths vanish. Physicists are free to add terms to the potentials that leave the fields invariant, a flexibility known as gauge freedom.

Maxwell's classical view, in which the potentials are not physical, was radically revised with the advent of quantum mechanics in the early 20th century. Heisenberg's uncertainty principle is incompatible with the notion of a point particle. Particles were replaced with wavefunctions $\psi(x) = |\psi(x)|e^{i\theta(x)}$, with an amplitude $\psi(x)$ and a phase $\theta(x)$. The squared amplitude $|\psi(x)|^2$ is the probability density of finding the particle at position x, whereas the phase is observed in interference experiments. For instance, the superposition of two waves with equal amplitude but opposite phase has vanishing amplitude and, hence, vanishing probability density. The phase of a particle is superfluous in classical mechanics, but it's crucial in quantum mechanics. That picture is analogous to the fields, not the potentials, being physical in electrodynamics.

Magnetic and electric effects

Those parallels between classical mechanics and electrodynamics are no coincidence. In quantum mechanics, the phase of a charged particle is fundamentally connected to the electromagnetic potentials—the magnetic potential **A** and electric potential ϕ. (See the article by Herman Batelaan and Akira Tonomura, PHYSICS TODAY, September 2009, page 38.) The phase of a particle with charge q traversing a spatial trajectory γ over a time interval τ changes by

$$\Delta\theta = \frac{q}{\hbar}\left(\int_\gamma d\mathbf{s}\cdot\mathbf{A} - \int_\tau dt'\phi\right), \tag{1}$$

relative to the phase of a particle that traverses the same trajectory with vanishing potentials. Because differences in phases can be observed in quantum interference experiments, the equation has consequences for the physical significance of the potentials **A** and ϕ.

Consider the case of two identical charged particles in a beam circling on opposite sides of a solenoid—a long straight tube tightly wound with coils of a current-carrying wire—illustrated in panel a of the figure. The current produces a magnetic field that is nonzero only inside the solenoid, and that field is related to the magnetic potential through its curl, $\mathbf{B} = \nabla \times \mathbf{A}$. When the particle wavefunctions interfere on the far side of the solenoid, a region of zero magnetic field, they still acquire a total phase difference of $\Delta\theta = q\Phi_B/\hbar$, where Φ_B is the magnetic flux in the solenoid. That so-called magnetic Aharonov–Bohm effect, observed experimentally in 1960, demonstrates that the magnetic potential must have a real effect that is absent from classical electromagnetism.

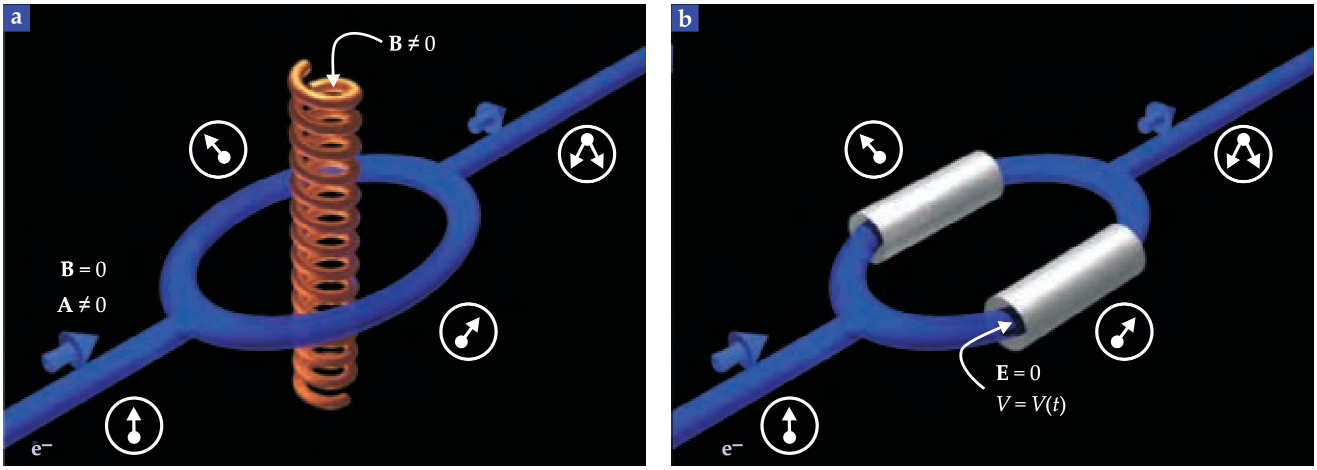
What about the electric Aharonov–Bohm effect? Consider the same two identical charged particles, but now pass them through two conductive tubes—so-called Faraday cages—illustrated in panel b. The objective of that setup is to induce an electric potential difference between the two particles without having them encounter any electric fields. To achieve that goal, a voltage ΔV is applied between the cages for a time ΔT while the particles are inside the tubes. The electric field is nonzero only outside the cages and only during the time the particles are inside them. According to the phase–voltage relationship, the particles acquire a relative phase shift

$$\Delta\theta \approx \frac{q}{\mathrm{e}} \times \frac{\Delta V}{\mu\mathrm{V}} \times \frac{\Delta T}{\mathrm{ns}}. \tag{2}$$

This equation describes the electric Aharonov–Bohm effect, in which the electric potential determines the evolution of the phase as a charged particle traverses time. The units of microvolts and nanoseconds are included to give a sense of the relevant scales in changing the electron's phase. An observation of a phase shift in the absence of electric fields acting on the particles would be a direct indication that the electric potential is itself physical. But to date no such observation has been made.

Experimental obstacles

To appreciate why the electric Aharonov–Bohm effect has not yet been observed, consider the units in equation 2. Exposing



MAGNETIC AND ELECTRIC Aharonov–Bohm experiments. **(a)** A copper solenoid produces a magnetic field inside it but prevents that field (and hence any Lorentz force) from existing outside it. Yet when a split electron beam (blue) passes around the solenoid, electrons in the different paths interfere and exhibit a phase difference—evidence of the magnetic potential's influence. The electrons' phases are drawn as white arrows, which evolve (or rotate) differently, depending on the local potentials, and produce a potential-dependent interference pattern (two arrows). **(b)** In the electric effect, the split electron beams pass through separate charged metallic tubes, inside of which an electric potential can exist but not an electric field. If an enclosed magnetic flux can cause phase shifts, one would expect an enclosed electric flux to do the same. White arrows again prescribe possibly interfering phases. (Images by Eve Vavagiakis and Donna Padian.)

an electron to a potential difference of 1 μV for 1 ns would change the electron's phase by roughly 2π, or one full cycle. That large magnitude makes the phase change challenging to observe experimentally. Even small fluctuations in the electric potential can blur out the interference pattern. A second, perhaps even more challenging issue is the elimination of electric fields. It is relatively simple to move electrons past a charged capacitor and observe the resulting voltage-dependent interference pattern—a tack used in earlier work—but very difficult to eliminate the effect of electric fields outside the capacitor, which makes the significance of the electric potential ambiguous. Yet a third obstacle to conducting a conclusive experiment is the interactions between free electrons and their environment, which cause their wavefunctions to collapse. Those interactions render the use of the electron wavefunction in an electric Aharonov–Bohm type of experiment extremely difficult. But what if a much bigger and more stable wavefunction could be used?

Fortunately, wavefunctions exist that are far better suited for an electric Aharonov–Bohm experiment. In superconductors, electrons below a critical temperature form a quantum condensate of electron pairs. The condensate has a phase that is experimentally measurable. And as macroscopic solids, superconductors are much less vulnerable to interactions that would perturb experiments on free electrons. Even so, the phase of the electron pairs in a superconductor is still sensitive to the electric potential via equation 2, which, when applied to superconductors, constitutes the famous second Josephson relation that helped earn Brian Josephson part of the 1973 Nobel Prize in Physics. (See PHYSICS TODAY, December 1973, page 73.)

Because equation 2 contains an electric potential difference, it applies regardless of whether any electric fields exist at the location of the superconductors. To our knowledge, the experiments that measured the phase shift have done so only with electric fields applied to the superconductors, so they were not able to test the electric Aharonov–Bohm effect. You need a region free of electric fields containing only an electric potential. Here's our proposal: Place a large planar capacitor between two small superconductors held close to its top and bottom plates. As the capacitor becomes charged, it induces a potential difference on the superconductors but only a vanishingly small electric field. That setup avoids the pitfalls mentioned above. The relative phase is routinely measured in superconductors. And because they are static, it should be less challenging to avoid exposing them to external electric fields or any other environmental interaction.

Demonstrating the physical significance of the electric potential in that way would fill a gap in both undergraduate-level electromagnetism and quantum mechanics courses. And the impact of such a measurement would inform our understanding of nature beyond the two subjects. The electric potential is intimately tied to gauge invariance, one of the core principles of our modern understanding of fundamental physics and deeply intertwined with the paradoxes regarding the unification of quantum mechanics and gravity. If experimentally confirmed, the electric Aharonov–Bohm effect would be a type of memory effect, as the phase of any charged particle delicately depends on the particles' entire history in time. At the moment, for instance, no one knows what happens to that information as a charged particle falls into a black hole.

Additional resources

▶ H. J. Bernstein, A. V. Phillips, "Fiber bundles and quantum theory," *Sci. Am.*, July 1981, p. 122.

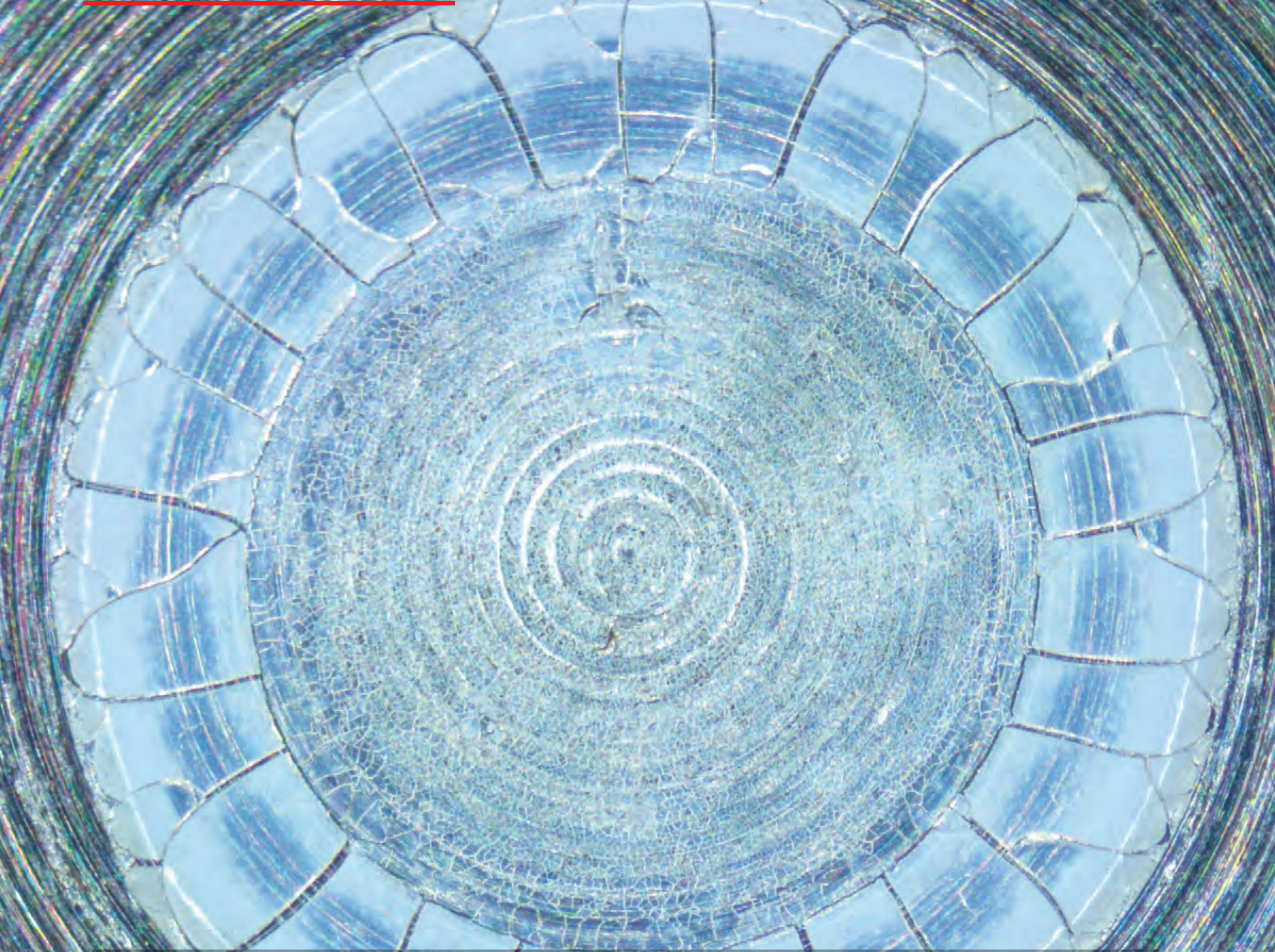
▶ R. G. Chambers, "Shift of an electron interference pattern by enclosed magnetic flux," *Phys. Rev. Lett.* **5**, 3 (1960).

▶ A. van Oudenaarden et al., "Magneto-electric Aharonov–Bohm effect in metal rings," *Nature* **391**, 768 (1998).

▶ T. C. Bachlechner, M. Kleban, "Proposal for testing the electric Aharonov–Bohm effect with superconductors," *Phys. Rev. B* **101**, 174504 (2020).

▶ L. Susskind, "Electromagnetic memory," https://arxiv.org/abs/1507.02584.

PT



Dynamics of drying

Cracks—fractures caused by material stress that exceeds the strength of the material—are ubiquitous in everyday life. They're also observed in the drying of silica suspension droplets, as seen in the microscopic image. Robert Mulka of Wrocław University of Science and Technology in Poland; Matthias Buschmann at the Institute of Air Handling and Refrigeration in Dresden, Germany; and their colleagues wanted to study cracking and the process of exsiccation, or drying out, in more detail. To do so, they dosed droplets of water-based suspensions with silica nanoparticles on a stainless steel substrate. A microscope with an attached camera then recorded the drying process in a laboratory environment with a constant temperature and relative humidity. This image, captured at the very end of the drying process, shows the characteristic "coffee ring" drying pattern.

A complex interplay of Marangoni convection and capillary flow inside the droplet transports the silica nanoparticles outwardly. As exsiccation progresses, the emerging tensile and shear stresses inside the deposition are relieved by the formation over time of a characteristic pattern of consecutive cracks. Each set of cracks has a specific morphology, which is linked to the previous stress distribution. The photo shows two morphologies. The tangential spiral cracks at the ring's edge develop first in random locations as bits of silica debris delaminate from the substrate. The straight, radially oriented cracks develop next as a response to the capillary pressure and shear stress operating between the droplet coating and substrate. (Submitted by Robert Mulka; R. Mulka et al., *Colloids Surf., A* **623**, 126730, 2021.)